Evaluation Framework for Water Quality Trading Programs in the Chesapeake Bay Watershed

May 29, 2009

This project was supported by the Chesapeake Bay Program Scientific and Technical Advisory Committee and the Mid-Atlantic Water Program. The Mid-Atlantic Water Program is supported by a grant from the USDA CSREES.

Authors:

Kurt Stephenson
Virginia Tech

Doug Parker
University of Maryland

Charles Abdalla
The Pennsylvania State University

Leonard Shabman
Resources for the Future

James Shortle
The Pennsylvania State University

Cy Jones
World Resources Institute

Bill Angstadt
Delaware Maryland Agribusiness Association

Dennis King
University of Maryland, Center for Environmental Science

Robert Rose
U.S. Environmental Protection Agency

David Hansen
University of Delaware

Evaluation Framework for Water Quality Trading Programs in the Chesapeake Bay Watershed

Water quality trading programs are being proposed and implemented across the US in a variety of forms and with differing objectives. The programs being proposed and implemented in the Chesapeake Bay region are no exception. Against this background the Chesapeake Bay Program's Scientific and Technical Advisory Committee and the Mid-Atlantic Water Program requested a general framework to inform and guide the evaluation of the performance trading programs. This resulting report was developed by a workgroup comprised of ten individuals with extensive experience in the study, design, and evaluation of trading programs. While the impetus for this report was to improve evaluation of trading programs in the Chesapeake Bay region, the evaluation framework is broad enough to apply to trading programs in general.

Different trading program designs reflect different program objectives. The objectives of a trading program will not only influence program design and implementation, but also the choice of criteria that should be used to measure performance. The first section describes the basic forms of trading, and in so doing illustrates the diverse objectives of trading programs.

The second section describes the use of *structural design criteria* as one approach to defining evaluation criteria. The structural criteria identify design requirements that may be reasonably predicted to achieve particular trading program objectives. This approach to evaluation assumes that, if implemented, programs that meet these design criteria will perform in a manner that is consistent with program goals.

The third section describes a set of *performance criteria* for evaluating trading programs. Performance criteria can be applied for *ex ante* or *ex post* evaluation (Tietenberg and Johnstone 2004). In an *ex ante* evaluation, such criteria are applied prior to program implementation to help select among alternative proposed designs. In an *ex post* evaluation, the criteria are used to evaluate the outcomes of a program that has been implemented. In the latter case, the existing trading program is evaluated against a select set of alternative water quality management instruments, including alternative trading designs that were not considered or, if considered, were not implemented. It is common to use the pre-implementation management strategy as the alternative in such an *ex post* assessment as it measures performance against what would have occurred had the pre-existing strategy been maintained. In this case, the basic question is whether performance with trading is better than what had existed before. But *ex post* assessments can also use as alternative strategies that might be of interest. Such an evaluation would help water quality managers evaluate whether and how alternative trading program designs, or alternatives to trading, would perform in comparison to a trading program that had been implemented.

Whether *ex ante* or *ex post*, evaluating programs based on performance criteria, can be a complex exercise because the outcomes of any trading program must be evaluated against alternatives that could have been (but have not been) implemented in lieu of the current program. This requires a clear conceptual model, description of the alternatives, and a comparison of a counter-factual situation that assumes how both regulators and regulated parties would have behaved under a different set of program rules and circumstances. This section concludes with an illustration of how evaluation of program performance can be simplified.

The report's final section contains conclusions that provide for both an understanding of the different forms of trading and the differences among programs. These conclusions should help increase the clarity and focus of public discussion about the merits of different trading forms, the strengths and limitations of current programs, and how improvements, if identified, might be implemented.

Water Quality Trading Programs: An Overview

Water quality trading programs can be designed and implemented in a variety of ways (Woodward and Keiser 2002; Stephenson, Shabman and Boyd 2005; Shabman and Stephenson 2007; Shortle and Horan 2001, 2008). For the purposes of this report, a general classification framework is used to identify five regulatory and trading program types: individual effluent standards (no trading) and four different types of trading programs (cap-and-allowance, cap-and-direct, credit sales, and offsite compliance credits).

Individual effluent standards, sometimes referred to as command-and-control, seek to secure water quality outcomes measured at the point of discharge. Under the Clean Water Act (CWA), individual effluent standards have been the primary regulatory tool used to control polluting discharges. These standards are applied to point-source dischargers of specific pollutants and are administered through the National Pollutant Discharge Elimination System (NPDES). The NPDES makes discharges into regulated waters conditional on having an NPDES discharge permit, which specifies the maximum allowable rates of pollutant discharges (e.g., the concentration of phosphorous in the flow from a wastewater treatment plant).

NPDES effluent limits are of two types. One is a technology-based effluent limit. This type is developed by regulators taking into account available treatment technologies and their affordability. Technology-based effluent limits (TBELs) are typically applied uniformly across sources or classes of sources (e.g. new versus existing sources). The CWA also instructs regulatory authorities to periodically review technologies and revise standards downward over time, called technology forcing or “ratcheting down.” In the event that TBELs for regulated point sources prove inadequate for receiving waters to meet ambient water quality standards, more stringent effluent limitations, called water quality based effluent limits (WQBELs) must be imposed by the permitting agency. Such a standard still may not secure ambient water quality standards when nonregulated discharges are the primary cause of the water quality standard violation.

To ensure compliance with effluent limitations, individual NPDES permittees must submit measures and records of effluent concentration and flow volumes according to procedures outlined by discharge monitoring reports (DMRs). In addition to numeric concentration and load limits, individual NPDES permits may also contain instructions and conditions on the use of installed effluent control technologies.

Traditionally, NPDES permittees were required to meet their effluent limits at the point of discharge. This remains the case for TBELs (US EPA 2007). But, the adoption of water quality trading programs can expand the options available for permittees to meet their WQBELs. Specifically, reductions in discharges from other point sources, or nonpoint sources if allowed, can be used to an NPDES permittee’s WQBEL.

From Individual Effluent Limits to Trading

Water quality-trading programs are not a substitute for government restrictions on effluent discharges. They are instead mechanisms for allocating discharges among alternative sources. More importantly under some designs a trading program can facilitate and motivate individual dischargers and groups of dischargers to identify and adopt innovative ways to reduce pollutant discharges. Trading programs allow discharge sources, to meet their regulatory requirements through effluent reductions made by other regulated or possibly unregulated, sources. The option to trade to meet effluent limits is a major source of the appeal of trading because it can allow for discharge allocations that reduce the costs of achieving water quality objectives over time and if designed in certain ways will accommodate economic growth while still maintaining water quality standards (Shabman and Stephenson 2007; Shortle and Horan 2008, Stephenson, Shabman and Boyd 2005)

Shifting effluent loads between sources can be accomplished in fundamentally different ways. Two key features – effluent cap participation and decision-making authority – can be used to broadly distinguish between types of trading programs. Each distinguishing feature of trading may be based on fundamentally different rationales and will have important implications for how trading programs are designed and implemented. These multiple trading program forms can also reflect multiple and sometimes distinct trading program objectives.

Effluent Cap Participation and Trading

Trading programs can first be distinguished by the scope of a load *cap*. A load cap is the collective allowable load that may be discharged by an identified group of dischargers. In turn, the cap is then apportioned to individual dischargers either through administrative allocation rules or through auctions. When the cap is set and then divided among those sources under the cap, the individual assignments are sometimes called discharge *allowances*[1]. A discharge allowance is the legal authorization to discharge a specific quantity of pollutant within a specified time period. Participating sources face legally binding obligations to hold allowances in order to lawfully discharge. The cap is the sum of the legal requirements to control discharge from an identified group of dischargers.

Failure to limit total discharges below the allowance holdings would trigger enforcement actions and the imposition of financial penalties against one or more sources under the cap. For purposes of this document caps are distinguished from load *targets*. The Chesapeake Bay Program established a 175 million load *target* for total nitrogen to the Bay from all sources (http://cap.chesapeakebay.net/progress.htm). The load target established a policy goal, but is not translated (completely) into legal obligations to limit discharges for all contributing sources.

One type of trading program allows for a system of reallocation of effluent load control responsibilities within (or under) a predefined effluent load cap. Such a trading program grants dischargers authority to transfer allowances to other dischargers subject to the cap. Allowances may be defined when the cap is set by allocating a fixed number of discharge authorizations to all the sources under the cap, with the sum of the authorizations equaling the cap. In this case the authorizations to discharge are able to be exchanged. A second approach is to focus on the NPDES effluent limits at each source. Over compliance generates a “credit” that can be then traded to a source that has under complied. Thus, what is traded between sources over and under compliance with NPDES effluent limits and the cap is the sum of the NPDES permit limits for the sources under the cap.

The logic here is the same as the way credits are generated for unregulated sources (see below for more discussion). In the cap case the baseline for credit generation is NPDES permit limit. Unlike allowances (which are defined in advance by a government agency), credits do not exist until reductions are undertaken and documented by the discharge source. In either case the trade will change the location of effluent discharges among sources subject to the cap (“point-to-point trading”).

[1] The legal permission to discharge a given amount of effluent is called a variety of names including wasteload allocation (WLA).

The second type of trading program involves a regulated source or group of regulated sources meeting its effluent limitation by securing effluent reductions from a source or sources outside the cap (by implication, a source not subject to legally binding effluent control limitations). In Chesapeake Bay region states, this situation universally applies to so-called trades between regulated point sources and unregulated nonpoint sources. The regulated sources face a mandatory requirement to limit discharges, but would be allowed to increase discharges above that limit by sponsoring equivalent load reductions at an unregulated source (typically thought to be an agricultural operation).

Reductions made by uncapped sources for sale to capped sources (ex. nonpoint sources) are commonly termed "credits". The load reductions eligible for trading by an uncapped source are calculated as the difference between a load with new effluent controls and a defined baseline. Unlike the credits created in the capped systems there is no regulatory load reduction requirement that can serve as the baseline.

Trading either within or outside a cap is an important consideration in trading program design because it has implications for the defined purposes of a trading program and the ability to achieve those purposes. In general trades within a cap provide the public with more assurances that water quality objectives will be secured. A capped program requires a comprehensive accounting and tracking of loads because regulatory responsibilities and load accounting is defined systematically in advance of program implementation. Total permissible loads are fixed and define by public agencies in advance of trades. Sources subject to a cap are required to limit discharges to an aggregate, socially sanctioned level in the face of economic and population growth. Because existing loads and allowances are defined in advance, any transfer of allowances between dischargers cannot, when properly executed, increase total loads.

Trades outside a cap face challenges of providing the same level of assurance that results from trades within a cap. In most trades outside a cap, the credit-creating seller is not pre-assigned a load baseline. Thus, some procedure must be devised to establish a baseline load. The procedure is only applied when a source decides to create a credit for trade. The concept of additionality requires the trading program to determine that a claimed reduction would not have been achieved in absence of a trade. Leakage is another issue that may be more challenging for trades with uncapped sources. Leakage occurs when a trade results in unexpected and unaccounted for net increases in loads (see discussion below under the general heading of "Water Quality Assurances").

Decision-Making Authority and Trading

Trading programs can also be distinguished by *who* has primary responsibility over the management of effluent waste streams. In discharger-directed trading programs, dischargers themselves have primary responsibility to determine how much they will discharge and how discharge will be managed. In a regulator-directed system, the regulator assumes primary responsibility for how loads will be managed.[2]

Discharger-directed trading programs are motivated by private financial incentives and facilitated by decentralized decision-making. Decentralized decision-making in general relies on individual dischargers to make decisions over what pollution prevention strategies to pursue, how to operate pollution control technologies, and how many discharge allowances to hold. The authority to make decisions about how to manage waste streams, however, is defined and bounded by a general set of socially sanctioned rules that establish basic rules and identify impermissible behaviors. For example, dischargers are required to hold allowances as a condition to legally discharge and are often prohibited from contributing to any localized water quality impairments. Discharger directed-trading programs under a cap with allowances are also called cap and allowance trading programs. A variant would be a cap and credit trading program.

Decentralized decision-making means that dischargers determine how many credits or allowances to buy and sell and under what terms of exchange. Under an allowance program design, the loads traded are fully at the discretion of the buyers and sellers, while in the credit trading design the trades are made in reference to seeking compliance with the technology based NPDES permit limit. To exercise whatever discretion is present, trading programs must provide relatively low cost mechanisms for buyers and sellers to find each other, agree on terms of exchange, enter into contracts and then be assured of enforcement of contract terms (transaction costs).

It is this program design where the financial incentives created by the opportunity to trade provide pollution prevention incentives (noted earlier). The opportunity to trade creates a financial consequence on the decision to discharge for all dischargers (prospective buyers and sellers). Prospective sellers have incentives to reduce discharges or costs in order to increase financial gains from trading. Prospective buyers seek to reduce costs and discharge of their own waste stream in order to reduce compliance costs.

---

[2] In any trading program, the responsibility for establishing the mass load cap and general trading rules is the exclusive responsibility of government. The issue confronted here is who has responsibility to make decisions about how to manage effluent discharge under a mass load cap.

Dischargers must also be granted the means to respond to these financial incentives. A discharger-based program grants discharge sources the authority to manage *how* pollutants will be controlled. A fundamental requirement of a cap and allowance or credit trading program is that it be performance-focused. Discharge-based trading programs focus attention on verifying the performance of dischargers (outcomes such as pounds of effluent released) and granting dischargers the authority to decide the means by which effluent will be controlled. Dischargers are allowed to investigate, experiment, and implement pollution control strategies and technologies that are best adapted to their individual circumstances in an allowance program: there may be less room for innovation in the credit based program. The combination of financial incentives and decision-making flexibility provides dischargers with both the reason and the means for developing and implementing possible new, low-cost ways to reduce loads.

Discharger-oriented trading programs, especially those based on allowances, are justified on the premise that knowledge of pollution prevention opportunities is fragmented, dispersed and incomplete. Those who advocate cap and allowance trading programs believe that the greatest potential for the development and application of new knowledge rests with those who stand to gain the most from its application and those with on-the-ground experience and knowledge of productive activities. In addition, the development and discovery of new knowledge is constrained by individual decision-makers' cognitive abilities. Unless given a reason to engage in the difficult and time-consuming process of thinking about wastewater management, many pollution prevention opportunities will simply go unrealized by both dischargers and regulators. With financial incentives, individual decision-makers who make pollutant generation, treatment, and discharge decisions are alerted to the importance of pollution prevention opportunities.

In contrast, a regulator-directed trading program assumes a centralized regulatory agency is capable of assigning and reallocating effluent control activity among sources. Regulator knowledge of the readily available control technologies and the site-specific costs associated with individual plants would allow such reallocations. Given the presumed knowledge of control options, a regulator directed program might define acceptable technologies or prescribe specific technologies.

In contrast to a discharger-oriented approach, the regulators also assume primary responsibility for ensuring that the cap is achieved through time. Economic growth can occur without violating the aggregate effluent cap through the "technology-forcing" actions of regulators. As new technologies are developed and proven, regulators may elect to lower the cap or impose more stringent technology requirements on selected subsets of dischargers. In this way, additional room is made for growth of new (both capped and uncapped) sources. Lowering aggregate discharge caps reduces the allowable discharges for sources under a cap and increases the demand for uncapped load reductions (application of pollution abatement controls outside a cap).

Regulator-directed trading is based on the premise that regulators, rather than dischargers, are in the best position to administer an effluent load management system. Such a trading program relies on regulatory authorities to direct or manage shifts in effluent loads between sources. Regulatory agencies are thought to be made up of pollution control experts with extensive experience and training in pollution control technologies. Publicly supported research grants are expected to target the most promising pollution control opportunities for development. Centralized expertise within regulatory agencies is deemed capable of developing systematic watershed scale plans to achieve water quality objectives. This position may be reinforced by a belief that dischargers are not as capable of identifying, or as willing to make, pollutant reduction investments.

These two distinguishing features of trading programs can be paired to create a taxonomy of trading types (Figure 1). Trades can occur in the context of a capped system on discharge allowances where both buyer and seller face mandatory requirements to hold allowances as a condition to discharge (top row). A discharger-directed capped program is *Cap-and-Allowance Trading.* (CAT) In a CAT program, individual transferable waste load allocations are called discharge *allowances*. CAT programs grant dischargers the discretion to manage loads and, therefore, allowances become the commodity exchanged between dischargers. As a variant of CAT and one that offers some (but less) discharger discretion, there is the cap and credit trading program, where the cap is derived from the sum of the individual NPDES permit limits. A capped, but regulator-directed trading program is called *Cap-and-Direct* (CAD). Such programs rely on regulators to allocate effluent control responsibility in order to achieve and maintain a cap. Similarly, trading programs involving trades between capped and uncapped sources can be distinguished by the authority to make effluent management decisions (bottom row). In these cases the capped sources are always the buyers of credits and the uncapped sources are always the sellers. In what is called *Credit Sales* (CS), the credit is a load reduction (often beyond a pre-determined baseline) achieved by the unregulated source. Under a credit sales program, sources of credits are granted the flexibility to decide how loads are controlled. The decision to buy credits and from what source is made by the capped source. Under an *Offsite Compliance Credit* (OCC) program, the regulator will establish how the credits will be produced by the unregulated source and the same regulator will determine when the capped source can sponsor load reductions at an uncapped source, as opposed to controlling the loads at their own location.

Whatever the form, there is another characteristic that will define the trading program: the exchange of legal responsibility for assuring that the required performance is achieved. In a cap and allowance program it is always the case that the entity that uses the allowance as the authorization to make a discharge is legally responsible for not making discharges that exceed the allowances held. However in the other systems it is possible that the legal responsibility for assuring that the load reduction occurs may transfer to the seller from the buyer, or may remain with the buyer. In the later case this term for the exchange will increase contracting costs.

Figure 1: Typology of Trading Programs

	Centrally Directed Decision-Making	**Decentralized Decision-Making**
Trade under a Cap (typically Point-to-Point)	Cap and Direct (CAD)	Cap and Allowance Trading (CAT)
Trade outside a Cap (typically Point-to-Nonpoint)	Offsite Compliance Credit (OCC)	Credit Sale (CS)

Trading Program Goals

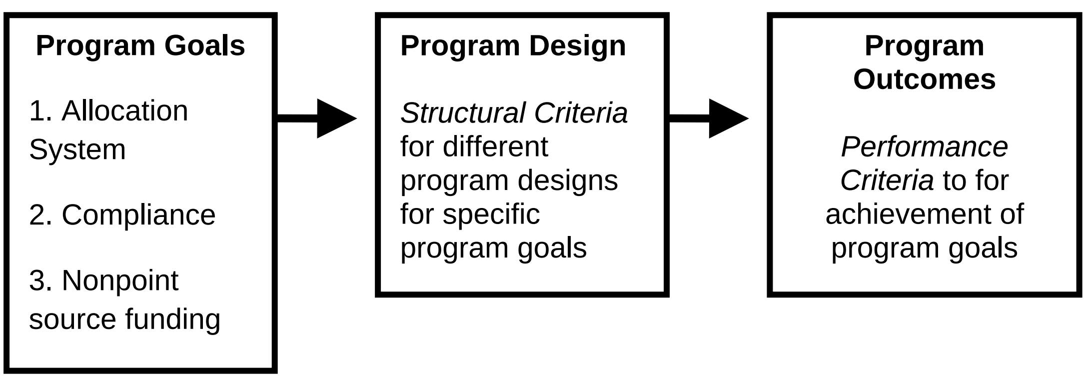
Trading programs can be designed to achieve different (and sometimes independent) program goals. A common goal of a trading program may be to develop an *allocation system* to achieve and maintain a socially prescribed aggregate load cap over time. If the cap is tied to achievement of water quality standards, a CAT program would maintain public water quality objectives over time. If there is no expected increase over time in loads from sources outside the cap the same result would be secured through a CAT program supplemented by a credit sale program. Trading of load responsibilities (allowances under a cap or credits outside a cap) is a means to this end, and not an end itself. Trading simply provides the incentives and opportunities for participating sources to effectively and economically comply with public water quality goals over time (Program Goal 1 in Figure 2).

One frequently adopted goal of a trading program is to create alternative compliance mechanism for an effluent standard (command and control) system (Program Goal 2 in Figure 2). As described above, for this form of program individual effluent limitations are set and sometimes lowered to limits-of-technology levels in order to meet water quality standards (water quality-based effluent limits). Regulated sources are expected to comply with the stringent standards. The water quality challenge, however, is confronted when economic and population growth increase wastewater flows and makes achievement of an individual mass load effluent limit technically infeasible. In such instances, regulatory agencies will require compliance with effluent concentration standards, but may allow or require regulated sources to meet individual effluent load limits by securing offsets and equivalent reductions offsite. The offsite effluent reduction activity (presumably securing reductions from a nonpoint source – a “trade”) becomes another permit condition for the regulated source.

Such programs do not create a system of allocating and reallocating load responsibilities under a mass load cap. This type of trading program is a way for regulatory agencies to maintain compliance with the traditional program of individual effluent limitations. It can be argued that such offsite effluent control responsibilities should not be called a "trade" at all. The source is not allowed to increase its discharges if it secures (presumably lower cost) discharges from another source. Indeed the source is directed by the regulatory authority to secure such offsets after it has met stringent effluent concentration limits in its wastewater that are chosen by the regulatory authority. The offsite control requirements are additional permit conditions typically set to meet unavoidable growths in flows. This type of trading program goal is a key component of the Maryland point-source trading program that was released in 2008.

Finally, funding for nonpoint source controls often appears to be an important goal of many trading programs (Program Goal 3 in Figure 2). In many parts of the U.S., the single largest cause of water quality impairments is effluent loads from unregulated (nonpoint) sources. Absent setting legally enforceable limits on these sources, payments (public cost share programs) are made to subsidize activities that are predicted to reduce effluent or to create a profit incentive for demonstrated pollutant reductions. Public funding from general tax revenues for these kinds of payment programs is insufficient to induce reductions to achieve water quality standards. In such cases, trading programs are often viewed, at least in part, as another source of revenue to finance reductions in nonpoint source loads. The generation of funds for nonpoint source load reductions appears to be an important program goal of the Pennsylvania nutrient trading program that was developed from 2005-2007.

Figure 2: Water Quality Trading Program Evaluation Framework



The various program goals do not necessarily overlap. For example, a trading program designed to achieve and maintain an effluent load cap may achieve that goal without ever leveraging resources for purchasing reductions from nonpoint sources. Similarly, the goal of maintaining compliance with individual effluent limitations at individual sources is mutually exclusive from the goal of designing a decentralized load allocation system. Understanding and acknowledging these different program goals is critical to program evaluation. Evaluating whether one program design is better than another leads to the next question, “better at doing what?”

Trading Program Evaluation

The trading program evaluation framework developed here can accommodate the evaluation of trading programs that exhibit variations in program goals and design types. Trading programs begin with general programmatic goals (Figure 2). Each goal can result in different trading program designs (as generally described in Figure 1). Program design describes the general structure and operational rules believed necessary to accomplish and satisfy trading program goals. Each trading program design will produce different behaviors on the part of dischargers and regulators, which in turn could conceivably produce different outcomes. Outcomes of interest may include changes in water quality conditions, the total quantity of effluent load discharged, the development of pollution prevention strategies, the cost of managing effluent loads, and administrative costs.

This section describes a set of general criteria that could be used in trading program evaluation. Two sets of program evaluative criteria are described. First, the *structural criteria* that could be used to evaluate program design are described. Structural criteria evaluation involves identification of a set of requirements necessary for a particular trading program type (Figure 1). For illustrative purposes, the structural criteria for discharger-oriented trading programs will be described below. Evaluating trading programs based solely on structural design assumes that there is a general agreement on program goals, dischargers’ and regulators’ behavior, and the performance that would be achieved if these structural criteria were met. While these indicators do not relate directly to outcomes, the advantage is that indicators of structural criteria are easier to measure and evaluate than performance criteria (see discussion below).

Next, a generalized set of *performance criteria* that might be used to evaluate economic and environmental outcomes are described. Other more specialized criteria that might be unique to specific program objectives are also identified. The discussion will include identification of the requirements and challenges of producing measurable performance indicators.

Structural Criteria

This section describes the structural criteria used to evaluate the design of cap and allowance trading and credit sale programs. Decentralized trading programs would, therefore, be evaluated by assessing how well program design matches these requirements. While some of these design criteria would be applicable to any trading program design, this discussion focuses on criteria for decentralized, performance-based trading programs

Within the context of water quality trading programs, structural criteria might be grouped into two general categories: choice criteria and water quality assurance criteria. Choice criteria describe the necessary conditions for dischargers to effectively participate in a trading program. The water quality assurance criteria is the set of requirements that ensure that water quality goals are achieved and maintained in the face of decentralized decision making over the type and location of effluent controls. Water quality criteria also broadly define the realm of responsibility and concern of regulatory agencies while choice criteria address the realm of discharger decision-making.

Table 1: Decentralized System Structural Design Criteria

Choice Criteria

Commodity Criteria
- Commodity defined in advance as effluent output
- Discretion to determine how effluent discharge is managed

Exchange Criteria
- Flexibility to determine with whom and when to trade
- Allow/facilitate multiple trade opportunities
- Reduce transaction costs of transfers
- Convey information about prices and trade opportunities
- Define rules and terms of trade in advance

Investment Environment Criteria
- Rule stability

Water Quality Assurance Criteria

Cap Setting Criteria
- Inclusiveness (new and existing sources)
- Size of cap

Equivalence Criteria
- Equivalence
- Load Accounting
    - Additionality
    - Leakage

Enforcement/Accountability Criteria
- Monitor
- Penalty

Choice Criteria

Choice criteria include criteria for defining the transferable commodity, the system of exchange, and investment environment.

Commodity Criteria

All trading programs require a standardized commodity (or commodities). In water quality trading, the commodity is usually defined in terms of a type of effluent discharge expressed in terms of mass load and defined over a period of time (e.g. pounds per year) and is reflected in an authorization to discharge a certain amount of pollutant. The commodity in trading programs is defined by the program and to the extent practical, should be defined in advance so as to provide dischargers clear knowledge and certainty regarding the conditions of the discharge authorization and the availability of the commodity. For example, a discharge allowance as defined above is the authorization to discharge a fixed quantity of effluent within a specific period of time. This commodity is defined at the start of the program. The commodity must be held in sufficient quantities in order to legally discharge (see enforcement criteria below).

Credits, on the other hand, are created only when a discharger can claim (according to program rules) that total discharge is less than an identified baseline level of discharge (Shabman, Stephenson, and Shobe 2002). Defining a credit requires an approved credit-generating methodology including how loads are quantified and baselines are defined.

Dischargers holding or possessing allowances must also be granted discretion on how to manage their effluent discharge (Shortle and Horan 2008; Stephenson, Shabman, and Shobe 2002). While the allowance limits how much effluent can be discharged, program rules should strive to allow discretion on the selection of production technologies, pollution control equipment, or pollution prevention strategies that the discharger employs to limit waste discharges.[3] This discretion extends to all trade participants (buyers or sellers). This discretion also includes determining the extent to which effluent control will be managed within any given allowance holding.

Defining the commodity based on effluent discharge requires quantification of load. Given this emphasis, decentralized trading programs should be performance-based programs. Quantification of the effluent load discharged could be measured at the source. To the extent practical, the effluent output and effluent discharge should be directly measurable at a reasonable

---

[3] Assessing the degree of discretion should not be narrowly limited to just trading program rules, but should consider how complementary and supporting programs also expand or contract discretion. For example, NPDES permitting requirements or capital grant programs may lock dischargers into specific technologies or effluent discharge levels that restrict dischargers' ability to cost effectively meet their aggregate cap.

cost (Oates 1994). Multiple methods of load quantification are possible ranging from methods of continuous measurement of flow and effluent concentration to a variety of sampling protocols (examples include those in NPDES discharge monitoring reports). Loads can also be quantified through modeled estimates of practices, but to do so is to take a significant step back from a performance based program.[4]

Direct measurement of loads allows dischargers discretion to determine how to control discharge. The preference for direct measurement of effluent discharge, however, is conditioned on the premise that measurement accuracy and precision can be secured at a reasonable cost. In choosing a method to quantify loads a choice must be made in recognition of the expected measurement error (loss of accuracy or precision) and the costs required to increase measurement accuracy and precision. In cases where direct measurement of outcomes is deemed too costly for the benefit of directly measuring performance, trading program designers may elect to model outcomes from a defined list of acceptable practices or technologies (as is the current practice for agricultural best management practices). Limiting control options is a barrier to innovation, but also an understandable choice when contending with measurement cost/uncertainty. Thus trade-offs can exist between measurement costs (and certainty) and choice over pollutant control strategies. Trading programs can also be evaluated on the opportunity and incentives to expand choices and/or lower measurement costs. For instance, trading programs that quantify loads (allowances or credits) based on modeled loads for a defined list of technologies can offer mechanisms to expand those lists with new information or reward efforts to measure outcomes with more certainty (more directly) (Stephenson, Norris, Shabman 1998).

Exchange Criteria

Trading also requires a system of exchange. In general, an exchange system should provide participants multiple trade opportunities and discretion of when and with whom to trade. Exchange flexibility criteria would identify the extent to which dischargers decrease discharges, but also allow dischargers the discretion to increase discharges by purchasing additional discharge allowances (or credits).[5] Trade opportunities provide sources with the discretion to

---

[4] The performance-oriented criteria mean that trading programs strive to avoid tying allowance and credit definition too closely to a particular practice or technology. Restricting allowance or credit generation to a particular practice (BMP for example) limits choices and reduces incentives to seek and develop more effective, lower cost alternatives.

[5] Numerous examples of legal and regulatory barriers to exchange flexibility can exist. For instance, existing trading programs currently prohibit an individual point source discharger from every exceeding technology-based effluent limits (even if dischargers are reduced elsewhere). NPDES permit conditions are typically prescriptive and costly to change (Shabman, Stephenson, and Shobe 2002; Industrial Economics 2008) As a note of interest, US EPA's Water Quality Trading Policy foresees potentially allowing a collection of point sources to satisfy their

choose individual discharge levels in ways that are cost-effective. Exchange flexibility also includes assessing how many different trading options are available to program participants. Multiple trading opportunities are needed to create competition, which should lead to lower cost control options.

Effective exchange arrangements must also provide a low cost and timely way to evaluate and conduct potential trades. The cost of coordinating buying and selling between trading participants includes search/information costs, contracting costs (negotiating a trade), and trade approval costs. Effective exchange arrangements can also be evaluated on the effectiveness of delivering timely and accurate information about prices (bid and offer prices) and trade opportunities to trading participants.

It should be noted that there exists multiple exchange arrangements that might meet these general exchange criteria. For instance, a common perception is that formal market-like exchange arrangements are needed to convey bid and offer prices between buyers and sellers. Formal market exchange, however, may be not feasible given the limited size and scope of many water quality-trading programs. For example, a point source cap may include relatively few dischargers that intend to trade infrequently. Alternative arrangements, such as discharger-operated associations, can be used to meet the general criteria for an exchange system (low transaction cost among participants). Within the nonpoint credit context, electronic clearing houses, aggregators, or publicly operated credit resale banks are all different ways that could lower transaction costs between relatively few buyers and multiple small nonpoint source credit suppliers. The purpose of this discussion is not to advocate any particular exchange system, but rather to note that there might be multiple alternatives available that could conceivably be used to meet these general structural design criteria. Different exchange arrangements can be tailored to meet the unique circumstances of a particular trading program.

In water quality trading, exchange systems require defining terms of trade between participating sources. These transfers refer to the various trading ratios that are intended to assure equivalence in water quality outcomes. For example trading ratios are typically required to account for differential spatial impacts (sometimes called "attenuation ratios") and differences in load quantification between types of sources (explained in more detail below). The relevant point for a decision on evaluating choice criteria is that market-like trading programs operate best when terms of trade are defined in advance of trading activities and defined exclusively on assuring standardization of the unit of exchange (water quality equivalence).

---

collective effluent load limits defined by technology-based standards, via water quality trading. This change, however, would require EPA to revise its technology-based effluent guidelines. This would considerably expand exchange flexibility in water quality trading programs.

Investment Environment Criteria

Finally, decentralized trading programs require stable and well-defined rules. Such stability reduces unnecessary uncertainty in the rules surrounding the use and transfer of the transferable commodity (allowances and or credits). Well-defined trading program rules that are not subject to unanticipated change provide a supportive climate for engaging in trade activity and pursing pollution prevention investments. For example, rule changes that penalize beneficial and successful prior investment decisions also undermine incentives for dischargers to make future pollution prevention investments. This is a particular concern if pollution prevention innovations that reduce the cost and discharge of pollutants may induce regulatory authorities to lower allowable load limits.

Water Quality Assurance Criteria

Water quality assurance criteria are the second general class of structural criteria that must be met in cap-and-allowance trading/credit sales trading programs. Quality assurance is used here as the collective term for a broad range of trading program criteria that can help to ensure that the water quality objectives of a trading program are met.

Cap Setting Criteria

The establishment of the cap is a primary means by which trading programs achieve water quality standards. An important criterion for cap setting is inclusiveness of existing and new sources. Cap inclusiveness defines who must participate in the cap (who must legally be required to hold allowances and limit discharge). Typically, caps are applied to point sources. Cap coverage, however, should be as broad as feasible and encompass as many sources as possible.[6] Expanding cap coverage would not only include placing small point sources under a cap, but also nonpoint sources as well. Furthermore, requiring new and expanding sources to "buy-into" the existing cap by purchasing allowances also expands cap coverage. In other words, regulatory officials should not increase the size of the cap to accommodate new discharge sources. The more inclusive the cap, the greater assurance the public will have that water quality standards will be achieved and maintained.

The size of the cap refers to the total allowable pounds of effluent that may be legally released by sources subject to the cap. Conceptually, the size of the cap should be consistent with the water quality objective. A "TMDL-like" evaluation for determining the total load of a pollutant that will achieve the water quality standard at a point in the water body would be used to determine

---

[6] The feasibility of cap expansion could be limited by lack of legal mechanisms to enforce discharge requirements and the cost of identifying, quantifying, and enforcing mass load limits on different sizes or classes of discharge sources.

the cap. This analytical process would model not only the loads to the water body, but also the fate and transport of that load through the system in order to determine the effect of a load at one location on the water quality criterion in the target area. However, given the uncertainties in such TMDL evaluations (Shabman, et al. 2007) the trading program may need to anticipate and accommodate the possibility that load limits will be revised over time.

Criteria for establishing cap size becomes less clear when the cap only applies to a subset of sources (point sources) and the capped sources do not represent a large portion of the total effluent load entering the water body. Under these conditions it may be difficult to achieve water quality standards by capping a subset of sources. This is because the achievement of water quality goals also depends on effluent reductions achieved by non-capped sources in the watershed; hence it would not be possible for the trading parties alone to achieve the water quality goals. The appropriate size of the cap in such instances cannot be determined by objective criteria, but must be established with both water quality, cost and fairness consequences in mind.[7]

Equivalence Criteria

Trading programs must be able to translate spatial and source heterogeneity of pollutant loads into equivalent water quality results where the water quality standard is to be met, called equivalence. Decentralized trading programs require that responsibility for controlling effluent loads can be transferred within a defined watershed boundary. If the cap is based on an analysis for a particular spatial location of discharges there must be an assurance that any spatial reallocation will not result in a violation of the standard. Because trading programs are designed to address a water quality concern in a specific location (say an estuary), the spatial movement of loads across the watershed must account for fate and transport to the location of interest.

Responsibility for controlling loads might also be transferred between dischargers with different effluent discharge characteristics. Quantifying discharges for any particular source or classes of sources might have different degrees of certainty surrounding measurement of actual effluent load discharge. Water quality trading programs that involve regulated point sources acquiring credits from nonpoint sources must develop acceptable procedures for dealing with differences in uncertainties with the measurement or quantification of effluent load.

---

[7] The size of a partial cap (point source cap) can be established in different ways. One way is to establish the cap based on some proportional share of the total load target. For example, North Carolina has established 30% nutrient reduction goals for the Neuse and Tar-Pamlico basins and point source caps are based on a 30% reduction in aggregate point source loads from a reference year. Maryland and Virginia established point source caps based on a technology-oriented “bottom-up” approach. Rather than referencing the cap to the water quality objective, this approach applies a particular technology-based concentration standards on sources (in Bay states, this often approaches limits of technology requirements) and then aggregates upward into a load cap.

While a requirement of any ambient-based water quality management program, equivalence requires sufficient knowledge and analytical tools to assess existing pollutant loads and water quality conditions, and to estimate the impacts of changing discharge points in the watershed. It should be stressed, however, that such knowledge does not need to be perfect and complete. Any complex biological system will exhibit significant uncertainties that may not be irreducible. While the trading programs design focus on equivalence issues between point and nonpoint sources, equivalence in point source loads are also subject to unknowns and uncertainties (Stephenson, Norris and Shabman 1998). Thus the equivalence criteria are based on reasonable systems knowledge and acceptable levels of uncertainties associated with estimating equivalent loads across space and between sources.[8]

In addition to providing equivalency, a trading program must provide an adequate system of quantifying and accounting for changes in load. A sound system of load accounting ensures that trades do not occur at the expense of other pollutant load reductions in the watershed. A number of accounting issues might arise with trades outside a cap (point-nonpoint source trades). Additionality is defined as load reductions that occur as a result of a trade, but would not have occurred in absence of a trade. Additionality assures that net pollutant loading does not increase as a result of a trade. In point source-nonpoint source trading, violation of additionality could occur when a nonpoint source, such as a farm operation, has implemented a number of effluent control efforts. These efforts could have been initiated for a number of reasons unrelated to any possible trade. If the agricultural operation can then also claim these changes as a nonpoint source reduction credit and sell that credit to regulated source, no additional reductions were achieved to offset the higher point source load. Additionality arises because of the challenges of identifying an appropriate baseline from which to measure changes in effluent loads for sources without a requirement to limit loads and without a system to account for all loads.

Additionality should not be a concern with trading under a cap (point-point source trading). In this case, trading results in the same combined load for the trading partners that would have occurred in the absence of trading. While this situation might sometimes seem to violate the requirement that credits sold must be the result of pollutant load reductions that would not have occurred in the absence of the trading program, it does not endanger the watershed goals because it does not involve the use of load reductions that are needed to achieve those goals. Point source allowances are defined in advanced and fixed in supply. Hence, in allowance trading (point source—point source trading scenario) additionality is satisfied if the trading partners collectively achieve their combined allocations.

---

[8] Equivalence is not by any means unique to trading. Any ambient water quality management program that aims to manage multiple sources across the watershed must make management decisions that are based on translating different levels and distributions of effluent loads into equivalent water quality outcomes.

A closely related accounting problem called leakage occurs from incomplete load accounting. Leakage is the induced, but unaccounted for, increases in pollutant loadings that result from trade activity. For example, an agricultural operation could generate nonpoint credits by installing BMPs on a portion of its land, say by installing riparian buffers. Holding all other farming activities constant, the riparian buffer would reduce nutrient loads leaving the farm. But, the installation of forested buffers may take highly productive bottomland out of production, prompting the farmer to bring additional upland acres under active cultivation. If the intensified upland land use causes an increase in effluent loads that are not accounted for in the credit calculation, leakage occurs (see box below for other examples). The lack of regulatory controls over nonpoint sources beyond the riparian area makes this type of problem difficult to prevent.

Trading program designs can address the leakage issue in a number of ways. Procedures for whole entity accounting/reporting of load for all nonpoint source credit providers is one example of a program requirement to address leakage. Evaluation procedures to access or identify (either *ex post* or *ex ante*) the likelihood or magnitude of leakage may also be undertaken. In some cases, the opportunities and incentives for shifting loads may be relatively small. If relatively large, trading ratios might be amended to reflect the approximate magnitude of leakage that might expected to occur.[9]

While leakage raises design challenges for trades involving sources outside a cap (point-nonpoint), it should be pointed out that other types of leakage can occur. For example, cap leakage could also occur outside of the trading program. In partially capped systems, incentives may be created for potentially regulated parties to shift discharge from sources required to operate under a mass load cap to sources outside of a cap (see box below for examples of leakage). The opportunity for such load shifting outside of a cap stresses the importance of defining an expansive and comprehensive load cap.[10]

---

[9] As with any program design, the cost of implementing actions to address leakage issues should consider administrative cost and feasibility.

[10] This type of leakage is not unique to trading, but is an issue with any regulatory system aimed at limiting mass effluent loads.

Examples of Leakage

Leakage for Trades outside of Cap (e.g. point-nonpoint)

- Farm with livestock and cropland ceases application of animal manures on owned farm, but rents neighboring farm to apply animal manures; enrolls owned farm as generating credits, without enforcement of excess nutrients on rented farm.
- Farm with animals without cropland has been providing neighboring farms manure nutrients for application in compliance with a regulatory nutrient management plan; farm then contracts to generate credits by transporting manure to alternate use; but neighboring farms replace nutrient applications with commercial fertilizer. This results in no nutrient application reductions on the total aggregate farms.
- Dairy farmer offers a land use conversion from row crops to hay to generate credits, but to meet his dairy silage needs plows new bottomland elsewhere to plant corn.

Cap Leakage

- Stringent cap on point sources prompts local governments to approve more new developments not served by centralized public sewer system. Wastewater from new developments are served by small scale on-site treatment systems not subject to cap.

Enforcement/Accountability Criteria

All trading programs require systems of enforcement. Decentralized trading programs require monitoring and verification of discharge (performance), a system to track ownership of allowances or credits, and adequate penalties for noncompliance (discharges by a source in excess of that authorized under the program rules).

Within a capped system (ex. point-point trading under a cap) monitoring, tracking, and penalty system associated with ownership of allowances are defined in advance of any trades. Thus, the transfer of an allowance does not create any new enforcement activities. However, the legal and permitting mechanisms to register and enforce these trades differ across media. Within the CWA, changes in discharge control responsibility must occur within the NPDES permitting structure. All trading requirements will be incorporated into NPDES permits, either directly or indirectly by reference, and the existing CWA enforcement mechanisms can be applied to permit violations. The ability to satisfy other market-based criteria (exchange flexibility, cost) may necessitate that creative forms of permitting be implemented (Shabman and Stephenson 2007).

Trades outside of a cap (ex. point-nonpoint trades) must meet the same basic accountability criteria identified above, but differences in assuring accountability in credit creation and enforcement occur in the details of implementation. The differences occur for two primary reasons: credits are, by definition, only created in the event of a trade (not defined in advance) and the exchange involves at least one party that is not legally required to limit discharge in absence of a trade. The basic accountability of credit creation involves verification by the regulatory agency or an approved third party that the credits are actually being produced according to the credit generating protocols. Conceptually, the transfer of credits to a regulated source (point) would legally allow the buyer to increase discharges. However, the CWA does not provide any mechanisms for enforcement actions against unregulated nonpoint source credit suppliers. States would need to legislate or regulate such actions. The conditions of credit generation would typically become conditions in NPDES permits, thus nonperformance on the part of a credit supplier constitutes a permit violation by the buyer. New regulatory authority would have to be created as part of the trading rules, or alternatively, the contracts between buyers and sellers and existing contract law could be relied upon as the enforcement mechanism. In the former case, passage of new legislation or promulgation of new regulations may be required by the state. In the latter case, the trading rules would set minimum requirements for contract provisions but the state would have no direct role in enforcing the contracts.

Documenting compliance might also differ depending on how loads are quantified. For nonpoint source credits where loads are not directly measured, but quantified through modeled changes, compliance monitoring and verification would focus on a specific set of observable and verifiable actions/activities that are consistent with the modeled changes in load. Most water quality trading programs prescribe one type of nonpoint source quantification, usually best management practice estimates and modeling.[11]

The means to address the uncertainty of assuring equivalent water quality results and legal challenges from shifting load responsibilities to sources outside a cap have important implications for the choice criteria discussed earlier. For dischargers to effectively participate in trading programs, participating sources' contract and legal risks to exchange activities should be minimized (investment environment criteria). The main risk perceived by buyers is that the credits may not actually be generated, resulting in a default on the contract terms. Credit sellers may fear the same possibility—that events beyond their control could make it impossible to supply the credits necessary to fulfill the contract requirements.

[11] Although technical in nature and beyond the scope of this paper, ambient monitoring of water quality would better assure the public that nonpoint source load reductions are real. Quantification procedures could be developed to smooth or average yearly variations in ambient measurements. Another approach is to calculate a mass balance that quantifies total amounts of nitrogen, for example, that enters and leaves the farm.

A number of mechanisms can be used to reduce or eliminate the risk of default to the trading partners or to assign responsibilities in the event of contract defaults. Among them are:

- insurance credit pools;
- reconciliation periods after the end of the permit reporting period during which regulated dischargers could acquire additional credits if necessary;
- contract provisions that protect the buyer financially in the event of a default; and

state-imposed penalties for failure to supply credits;

Performance Criteria

The above discussion describes the basic structural criteria needed for a decentralized trading program. Structural criteria allow program evaluators to measure whether the conditions and requirements for a particularly trading program type are adequately implemented. Structural criteria, however, do not directly evaluate program outcomes.

This section describes performance criteria that can be used to evaluate trading program outcomes. The section begins by identifying a common and general list of performance criteria. Criteria may be measurable either by quantitative assessment of collected data or qualitative assessment based on expert consensus judgment. In actual program evaluation, the choice of performance criteria will depend on program goals of the specific trading program. Examples will be given of both general performance criteria and program specific performance criteria. The section concludes with a brief discussion of the requirements and challenges of evaluating trading program performance.

General Performance Criteria

A general set of water quality performance criteria are: 1) achievement of water quality objectives; 2) cost effectiveness; 3) pollution prevention innovation; and 4) equity/fairness (Tietenberg and Johnstone 2004; Shortle and Horan 2008) (see Table 2).

Trading programs are implemented primarily to assist in the achievement of water quality objectives. A variety of performance criteria and indicators can be used to evaluate performance of meeting water quality objectives (see Table 2). Some of these water quality trading program performance criteria are similar to the structural criteria discussed above.[12]

---

[12] Note that many trading program supporters promote the ancillary benefits from trading (i.e., stream shading and wildlife habitat from riparian buffers). While these benefits may receive a lot of attention, they are not primary to the goals of the program. Thus, we do not include these benefits in our evaluation framework.

Table 2: General Program Performance Criteria

Achievement of Water Quality Objectives
- Watershed Level Criteria
  - Percentage of total load covered by the mass load cap
  - Compliance with mass load cap
  - Cap leakage
  - Avoidance of unintended localized water quality impairments ("hot spots")
- Trade Accounting
  - Adequate baselines for nonpoint source credits
  - Evidence of leakage
- Accountability/Enforcement
  - Monitoring/measurement /verification of outcomes
  - Penalties for noncompliance

Cost Effectiveness
- Entity-level cost effectiveness
- Allocative cost effectiveness
- Public administration costs
- Private administrative/transaction costs

Pollution Prevention Innovation
- Investments in research and development
- Development of patents in pollutant control technologies
- Improvement in pollutant removal efficiencies over time
- Reduction in pollutant control costs over time

Equity/Fairness
- Distribution of load responsibility between classes of sources
- Distribution of allowances to regulated (point) sources under a cap
- Distribution of costs among sources and between classes of sources
- Distribution of public subsidies between sources and classes of sources

Achievement of Water Quality Objectives

At a watershed level, a mass load cap represents a legally binding requirement to limit discharges to levels consistent with water quality standards. The greater the portion of discharge sources that face no mandatory limits, the less like water quality goals will be achieved and maintained over time. Thus an important criterion for water quality trading programs is the portion of total effluent load subject to a mass load cap.

In cases where partial caps apply and capped sources constitute only a relatively small portion of the overall load, it would be unrealistic to expect a trading program alone to achieve the overall water quality objectives for the entire watershed. In this case, an appropriate criterion would be achievement and maintenance (long-term compliance) of the mass load cap on the covered sources over time. Finally, a watershed level criterion would be the absence of localized water quality impairments due to trading activity (avoidance of the creation of “hot” spots).

Criteria to identify consistency with water quality goals can also be identified at the scale of individual trades. For trades with sources without quantified load limits (point-nonpoint trades) water quality criteria can be identified to determine whether claimed credit reductions actually occur. Criteria for ensuring there is additionality include clear baseline requirements. Performance criteria would also identify the potential and occurrence of leakage.

Finally program outcomes can be assessed based on criteria that ensure adequate compliance monitoring, verification, and enforcement. This includes assessment of active monitoring and enforcement activities to document performance and compliance with trading program rules.

Cost Effectiveness

A key factor motivating interest in trading is the expectation that trading programs can achieve water quality goals at lower social cost that alternative mechanisms. Accordingly, a second set of criteria to evaluate trading programs relates to the cost of achieving water quality objectives. A program is cost effective when it minimizes the full implementation cost of achieving a specific, pre-defined objective.[13]

The cost to polluters of a pollution control allocation is the aggregation of their individual costs. Their entity-level costs will be a function of the combination of inputs, technologies, and practices that they select to achieve their individual effluent limits. Cost-effectiveness is served by providing polluters with discretion to choose the set of inputs, technologies, and practices that allows them to minimize their individual costs. However, entity-level cost efficiency does not guarantee system-level cost effectiveness. Each discharge source could individually minimize the cost of meeting individual discharge goal, but the overall allocation of control responsibility across sources might not minimize total system costs. Allocative cost-effectiveness requires that pollutant control responsibility across discharge sources be allocated in such a way as to minimize the sum of all pollutant control costs.

[13] Costs include both capital and operation costs, forgone returns on productive activities, transaction costs and public administration costs. Payments made from a trade and payment of public cost share should not be viewed as separate cost categories. These payments represent transfers between different parties and involve question of *who pays* the costs.

"Number of Trades" as an Indicator of Program Cost Effectiveness

The number of trades made in a trading program is a sometimes suggested indicator of cost effectiveness (EPA 2004). In some trading programs, a trade represents a voluntary exchange whereby a low cost source voluntarily undertakes greater responsibility for controlling pollutants while a higher cost source spends less on control. Such a trade represents a cost effective reallocation of effluent control responsibilities by shifting control responsibility to the low cost source.

As an indicator of program success, the number of trades should be interpreted with caution for two reasons. First, well-designed trading programs may stimulate rapid search, development, and implementation of previously unknown cost reducing measures at the entity-level (see pollution prevention indicators below). Trades and reallocations may not be initially forthcoming because of search for these entity-level cost efficiencies. Second, in some types of programs trades are neither voluntary nor done to reduce costs. As described in the first section, programs designed to maintain compliance with individual effluent standards may require trades as a last resort to meet effluent load limits. In such instances, trades are allowed only after the implementation of high cost expensive limits-of-control have been imposed and no additional control options are available. In such cases, trades are an indication that high costs have been reached, not avoided.

On the other hand, lack of trades due to high transaction costs, limited trade opportunities, or regulatory barriers would indeed be an impediment to cost effectiveness.

In addition to incurring costs from investments in equipment and changes in operations to reduce effluent loads, participants in trading will incur various forms of transactions costs from executing trades. High transactions costs can depress participation in trading and trade volumes, leading to participation and coordination failures that limit efficiency gains (Shortle and Horan 2008; Nguyen and Shortle 2007). A key task of agencies setting up trading programs is to consider the implications of trading rules on transactions costs, and to develop information and exchange mechanisms that foster low transactions costs.

Trading programs also entail public sector administration costs. Public agency costs include design, administration, trade approval, and additional monitoring/verification and enforcement costs beyond what would be required under a no-trade option. Trading program may also impose new coordination costs on regulated entities. Such coordination costs involve information, negotiation, permitting/contracting, and trade approval costs associated with reallocation and/or trading activity. As discussed earlier, multiple organizational, permitting, and market arrangements are available to manage and reduce these costs.

Pollution Prevention Innovation

A third evaluative criterion for water quality programs is the rate of pollutant control innovation. If water quality is to be maintained over time in the face of population and economic growth, it is necessary to continuously improve the technical efficiency of the prevention and treatment of pollution. Innovation also reduces scarce society resources needed to achieve public water quality goals. Different program designs create different pollution prevention incentives which could, in turn, produce differential rates of technical change.

Producing measurable indicators of pollution control innovation is challenging. Some indirect indicators could include the investments in research and development activities or the number of pollution control patents filed for a particular category of discharge sources. Yet, the expenditure of money on research and development does not necessarily translate into effective pollution control improvements. Similarly, technical improvements may occur due to operational and process refinements for specific production and pollution control technologies. These changes might be measured as improvements in effluent removal efficiencies or change in pollutant control costs over time.[14]

Equity/Fairness

Another evaluative criterion that may be difficult to define precisely but nonetheless plays an important role in comparing program outcomes is equity. Equity describes the perceived fairness of the distribution of program results and outcomes. A trading program may achieve water quality objectives and produce cost effective outcomes, but nonetheless produce a distribution of costs and outcomes deemed unfair from an equity perspective. Obviously, perceptions of program equity may differ not only across different types of trading programs, but also between different groups (point sources, agricultural sources, taxpayers, consumers, utility rate payers, etc).

While perception of fairness is a difficult concept to measure, numerous indicators can be used to describe the distribution of program outcomes. For example, the assignment and distribution of pollutant control responsibility across classes of sources will produce different costs across

---

[14] Identifying the effect of a trading program has on rates of innovation is analytically challenging (see discussion below). For these reasons, program evaluation might rely on structural evaluative criteria that measure the program's opportunities and incentives to implement new technologies.

dischargers. This distribution can be measured as the share of regulatory obligation to the contribution of total effluent load to the water body. Given the structure of the CWA, incentives exist for regulatory authorities to assign a larger burden of control responsibility to point sources that are already subject to NPDES permitting and/or are "easy" to identify and regulate. Regulated point sources may view trading programs as unfair if point sources bear the primary burden of achieving water quality standards when their total load is a small portion of the watershed total. Trading programs may add to this perception of unfairness if trading is perceived as a way to help finance controls that the public agencies refuse to directly regulate. Fairness concerns might also arise concerning the distribution of allowances (wasteload allocation) between regulated sources under a cap or the distribution of public cost-share funds (subsidies) among sources or classes of sources.

Program Specific Performance Criteria

The criteria just described are general standards with which a trading program will be evaluated. In an evaluation of specific programs, other criteria that relate to specific and unique program goals might be need to be identified. For instance, as discussed earlier, an important program goal of some trading programs seems to be the generation of funding/financing for implementing nonpoint source controls. If generating resources for purchasing nonpoint source reductions is a program end itself (rather than a means to an end), then the number and value of nonpoint source credit purchases annually becomes a performance metric. This performance criteria, however, only applies for trading with nonpoint source revenue generation as an explicit goal. Trading programs aimed at achieving a water quality objective for a group of dischargers can be highly successful based on the general set of performance criteria described above and never generate a single point-nonpoint source trade.

Evaluating Program Performance

The ultimate reason to undertake program evaluation is to identify if a trading program is successful in meeting these performance criteria. Evaluation of program performance requires an evaluation of program outcomes against a reference condition or alternative (Tietenberg and Johnstone 2004). Conceptually, program outcomes should be compared against outcomes that would have occurred if a program without trading or another trading program design had been adopted ("with versus without" the program).

For example, the general set of performance criteria could be applied to the four prototype trading programs described in Section 1. The reference condition would be the individual effluent standards implemented through NPDES permits and the nonpoint source reductions secured through voluntary nonpoint source cost-share programs aimed at implementing specific best management practices. A trading program evaluation would compare the outcomes achieved under the reference condition with those that might be achieved under different trading program designs. Table 3 shows a simple program evaluation matrix based on example evaluation. Of course, the table content would change depending on specific program goals and trading program designs.

While conceptually a "with vs. without" is the appropriate evaluation perspective, such an evaluation requires the construction of counter-factual outcomes of what would have occurred under a different program. Thus, program evaluation requires comparison of "alternative futures" based on different water quality management structures. The comparison of programs based on performance criteria is an inexact and analytically challenging (and costly) exercise. Whether allocative cost-efficiency is being achieved, whether dischargers are implementing lowest possible cost control options, and new technologies are being developed are all examples of criteria that must be compared against the predicted alternative outcomes (different trading program designs). Similarly, the degree to which water quality goals are achieved must be compared to the probability of achieving goals under an alternative policy. This requires the evaluator to either 1) use models to simulate behavior under different program designs or 2) perform a comparison of outcomes across different programs with similar background characteristics.

Constructing these counter-factual program outcomes can be done in a number of ways, including formal modeling, carefully structured empirical comparative system case studies, and expert judgment. Each method has limitations and therefore the results obtained have caveats. Computational models can estimate outcomes related to cost and discharge outcomes, but such models typically are based on simplified behavioral premises and assumptions that incompletely reflect different incentives, behavior, and outcomes that might arise across program designs. Comparative analysis attempts to construct plausible counter-factuals by examining outcomes across different existing regulatory programs that have similar initial water quality and socio-economic conditions. Experimental studies can compare behavior and outcomes under carefully controlled settings, but the experimental design, by definition, limits evaluation to specific, narrowly defined issues. Expert judgment involves the qualitative assessment of performance criteria from individuals that have extensive knowledge of the structure and operation of water quality management programs.

Simplifying Program Evaluation

Given the analytical challenges of program evaluation, procedures can be devised to simplify the performance evaluation of trading programs. The simplest form of program evaluation compares a pre-existing program with a program that includes some form of trading. The comparison is simply "before and after" and is not considered against possible counterfactual trading or non trading programs. Such an approach simply asks whether trading programs represent an improvement over what existed prior to program implementation.

Evaluation criteria are also simplified, and possibly, narrowed based on specific objectives of the trading program.   Indicators of expected/or realized outcomes after trading are compared to the outcomes before the particular trading design. This more limited type of evaluation can help program administrators decide whether the existing programs are an improvement from the former regulatory approach using their own internal criteria.

A simplified approach would be to use evaluative criteria that reflect the goals and objectives of specific trading programs. Evaluative criteria and measurable indicators will be developed that reflect these program objectives. When appropriate, the criteria described above would be used in this simplified program evaluation. The objective of this analysis is to evaluate trading programs against their own programmatic goals.

Conclusions and Recommendations

This main goal of this report was to describe a general framework that will help inform and guide future evaluation of performance of water quality trading programs. While the impetus for this report was a desire to evaluate trading programs in the Chesapeake Bay region, the evaluation framework is broadly applicable to trading programs in general.

The following conclusions and related recommendations can be drawn from this report:

The trading programs that have been proposed or implemented to address water quality in the Chesapeake Bay region and elsewhere in the U.S. are quite diverse. The particular program designs that have been chosen reflect different program objectives. The trading program's goal(s) not only influences program design and implementation, but will affect program performance and the choice of criteria that will be used to measure performance. ***Therefore***, the Chesapeake Bay Program should request that each trading program should review and clarify its explicit goals.

Despite the data limitations and difficulties, the evaluation approaches described here can illuminate, and if needed, improve the performance of trading programs for the future. ***Therefore***, The Chesapeake Bay Program should request that assessments of state water quality trading programs or experiments in the Chesapeake Bay should apply the systematic evaluation procedures described in this report.

Structural design criteria provide a more tractable approach for evaluating the performance of water quality trading programs. ***Therefore***, given the relative analytical difficulty of performance evaluation, evaluation using design criteria should be the initial effort in evaluation by the state government decision-makers responsible for each current program. The results for that evaluation should be available to stakeholders and member agencies of the Chesapeake Bay Program and can be used to make design adjustments as warranted to increase the likelihood that program goals will be achieved.

Data and analytical requirements for performance evaluation can be significant; however such evaluations may be desired as trading programs mature. ***Therefore***, as an initial step to developing a performance based assessment, The Chesapeake Bay Program should work with the states to develop an operational framework for performance evaluation should be developed, data and analytical requirements identified, and the costs and benefits for such an assessment evaluated.

References

Environmental Protection Agency. 2004. *Water Quality Trading Assessment Handbook: Can Water Quality Trading Advance Your Watershed Goals?* EPA publication EPA 841-B004-001, Washington DC.

Environmental Protection Agency, 2007. *Water Quality trading Toolkit for Permit Writers*. EPA 833-R-07-004, Office of Wastewater Management. Washington DC.

Horan, R., and J. Shortle. 2001. *Environmental Instruments for Agriculture*. In J. Shortle and D. Abler (Eds.), Environmental Policies for Agricultural Pollution. Wallingford, UK: CAB International.

Industrial Economics, 2008. *EPA Water Quality Trading Evaluation: Final Report.* U.S. Environmental Protection Agency, Washington DC.

Oates, W. 1994. *Can We Protect the Environment and Improve the Tax System at the Same Time?* Southern Economic Journal. 915-922.

Shabman, L., K. Stephenson, and W. Shobe. 2002. "Trading Programs for Environmental Management: Reflections on the Air and Water Experiences." *Environmental Practice* 4 (3): 153-162.

Shabman, L. and K. Stephenson. 2007. "Achieving Nutrient Water Quality Goals: Bringing Market-like Principles to Water Quality Management." *Journal of American Water Resources Association* 43 (4): 1076-1089.

Shabman, L., K. Reckhow, M. B. Beck, J. Benaman, S. Chapra, P. Freedman, M. Nellor, J. Rudek, D. Schwer, T. Stiles, C. Stow. 2007. *Adaptive Implementation of Water Quality Improvement Plans: Opportunities and Challenges.* Nicholas Institute for Environmental and Policy Solutions. NI R 07-03, Duke University, Durham North Carolina.

Shortle and R. D. Horan. 2001. The Economics of Nonpoint Pollution Control. *Journal of Economic* Surveys 15 (3): 255-289

Shortle, J.S. and R. D. Horan 2008. *The Economics of Water Quality Trading*. International Review of Environmental and Resource Economics: 2(2): 101-133.

Stephenson, K., L. Shabman and J. Boyd. 2005. "Taxonomy of Trading Programs: Concepts and Applications to TMDLs." In *Total Maximum Daily Load: Approaches and Challenges.* Ed, Tamim Younos, pp. 253-280. Tulsa OK: PennWell Press.

Stephenson, K., P. Norris, and L. Shabman. "Effluent Allowance Trading: The Nonpoint Source Challenge." *Contemporary Economic Policy* 16:4 (October 1998): 412 - 421.

Tietenberg, T. and N. Johnstone. 2004. "Ex Post Evaluation of Tradable Permits: Methodological Issues and Literature Review." in *Tradable Permits: Policy Evaluation*, *Design, and Reform*. Organization for Economic Co-operation and Development (OECD), Paris France.

Woodward, R. and R. A. Keiser. "Market Structures for U.S. Water Quality Trading" *Review of Agricultural Economics.* 24: 366-383

Table 3: Example of a Trading Program Evaluation Matrix

Performance Criteria	Individual Effluent Limits	Cap and Direct CAD	Cap-and-Allowance Markets CAM	Credit Sales CS	Offsite Compliance Credit OCC
Water Quality Watershed Level Criteria Trade Accounting Accountability/Enforcement					
Cost Effectiveness Entity-Level Allocative cost effectiveness Private Admin Costs Public Admin Costs					
Pollution Prevention R&D Investments Development of patents Improvement in removal efficiencies Reduction in control costs over time					
Equity Distribution of load responsibility Distribution of allowances Distribution of costs, subsidies					