Validation of the Modified Virginia Save-Our-Streams Protocol

J. Reese Voshell, Jr.
Stephen W. Hiner

Department of Entomology
Virginia Tech
Blacksburg, VA 24061

August 1, 2006

Introduction

Virginia Save-Our-Streams (VSOS) contracted Dr. Reese Voshell and Mr. Stephen Hiner to conduct a validation study of the current protocol used by this volunteer organization. The currently used protocol was developed by Engel and Voshell (2002) and was a modification of the original VSOS protocol (Firehock and West 1995). In this report, the currently used protocol will be referred to as the modified VSOS protocol. The multimetric index that is the output of the volunteer protocol will be referred to as the Virginia Save-Our-Streams Index (VSOSI).

The major purpose of this study was to determine if volunteers using the modified VSOS protocol reach the same decisions about the ecological condition of streams as formally trained aquatic biologists using professional methods. This proved to be true in the study by Engel and Voshell (2002), but the inherent variability in field studies requires that the modified VSOS protocol be further validated in other years and streams and with other volunteers. In addition, since the Engel and Voshell (2002) study the Virginia Department of Environmental Quality (VDEQ) has developed a new measure of stream health based on benthic macroinvertebrates, the Virginia Stream Condition Index (VSCI). This measure is now being used as the benchmark for whether streams are impaired or not and whether or not a TMDL is necessary. The emphasis of this study

was on the final output of the protocols for interpreting data and reaching conclusions, primarily the VSOSI determined by volunteers versus the VSCI determined by professional biologists.

Methods

This study was conducted in wadeable, rocky-bottomed sections of streams in the mountainous and western piedmont areas of Virginia. There were 21 study sites, which are listed in Table 1. At each site, VSOS certified volunteers and staff took a sample and reached a conclusion about ecological condition according to the modified VSOS protocol. Instead of returning the organisms to the stream alive as is usually done, the volunteers preserved and labeled each sample for later examination by the scientists conducting the validation study at Virginia Tech. At the beginning of the study, Virginia Department of Environmental Quality (VDEQ) staff trained VSOS staff in the field to collect the "professional" samples in accordance with the standard operating procedures used by that agency for riffle habitat. Thereafter, VSOS staff collected the professional samples at the same sites and at the same times as the volunteers. The professional samples were preserved in alcohol and delivered to the scientists conducting the validation study at Virginia Tech for sorting, identification, and data analysis according to standard operating procedures used by VDEQ.

The summary statistic for the volunteer samples is a multimetric index that aggregates the results of 6 individual metrics into a single score that can range from 0 to 12. In this report, this will be referred to as the Virginia Save-Our-Streams Index (VSOSI). For the VSOSI, scores of 7-12 are indicative of acceptable ecological conditions, whereas scores of 0-6 are indicative of unacceptable ecological conditions.

The summary statistic for professional samples taken by VDEQ professional biologists is a multimetric index named the Virginia Stream Condition Index (VSCI) that aggregates the results of 8 individual metrics into a single score that can range from 0 to 100. For the VSCI, scores of 60-100 are indicative of reference conditions (= acceptable conditions), and scores of 0-59 are indicative of impaired conditions (= unacceptable conditions). Although the focus of this study was a comparison of the VSOSI to the VSCI, another multimetric index was also calculated for the professional samples, the Macroinvertebrate Aggregated Index for Streams (MAIS). This was included because it was the only professional multimetric index that was available for the original study done by Engel and Voshell (2002), and, thus, it would be a good comparison for the new validation study. The MAIS aggregates the results of 9 individual metrics into a single score that can range from 0 to 18, with scores of 13-18 indicating acceptable ecological conditions and scores of 0-12 indicating unacceptable ecological conditions.

Three statistical analyses were performed to examine the agreement of the final outcomes of the volunteer and professional protocols. Pearson product-moment correlation analysis was used to test the relationship of the scores ($\propto = 0.05$). The desired outcome would be to have $p < 0.05$ (statistically significant) and r as high as possible but at least > 0.7. Having an r value > 0.7 means that at least 50% of the variation is explained by the relationship because $r^2 \geq 0.50$. In addition to having the VSOSI strongly related to the scores from the professional protocols, it is important that the VSOSI reach the same conclusions on the status of stream ecological condition (acceptable versus unacceptable). Chi-square goodness of fit and McNemar's test were used for this purpose (both with $\propto = 0.05$). For chi-square goodness of fit, the desired

outcome would be for the ratio of the number of sites classified as acceptable to the number of sites classified as unacceptable to be not different for the VSOSI and VSCI. For McNemar's test, it would be desirable for the same proportion of the same individual study sites to be classified as having unacceptable ecological conditions by both the VSOSI and VSCI. In addition to statistical analyses, the Virginia Tech investigators checked the taxonomic identifications of the macroinvertebrates made by volunteers in all 21 samples.

Results

The correlation of the VSOSI with the VSCI and MAIS at the 21 study sites is presented in Figure 1 (A and B, respectively). There was strong agreement between the VSOSI obtained by volunteers and both measures obtained by professional methods. The Pearson product moment correlation analysis indicated that more than 66% (r^2 for r value of 0.814) of the variance was explained by the relationship of the VSOSI and the VSCI. For the relationship of the VSOSI and MAIS, 54% (r^2 for r value of 0.735) of the variance was explained. Having at least 50% of the variance explained by the relationship of two variables indicates a strong relationship for biological field studies, which typically have a lot of unexplained inherent variability. The r^2 value of 0.66 for the correlation of the VSOSI and the VSCI is unusually high for biological field studies. The correlation of the VSOSI and the MAIS score was higher in this study than the previous study by Engel and Voshell (2002) in which the r^2 value was only 47.9%. The VSCI had not been developed at the time of the previous study.

The resulting ecological condition classification for the 21 study sites based upon the volunteer (VSOSI) and professional protocols (VSCI, MAIS) used at each site are

presented in Table 1. In spite of the strong correlation between the VSOSI and the professional indices shown in Figure 1, there were considerable differences in the classification of ecological condition reported in Table 1. The VSOSI classified 3 sites as unacceptable and 18 sites as acceptable; the VSCI classified 11 sites as impaired and 10 sites as reference; the MAIS classified 7 sites as unacceptable and 14 sites as acceptable.

One would hope that the classification of the 21 sites as acceptable or unacceptable by professional methods and volunteer methods would be in the same ratio. Chi-square goodness of fit was used to test the ratios (Zar 1999), with the *null hypothesis* that the ratios of unacceptable to acceptable ecological conditions were not different ($\propto$ = 0.05). For this test, the numbers of unacceptable and acceptable determinations of ecological condition at the 21 sites by means of the professional methods (VSCI, MAIS) were the expected frequency. The numbers of unacceptable versus acceptable determinations of ecological condition at the 21 sites by means of the volunteer method (VSOSI) were the observed frequency. Results of the chi-square goodness of fit test are presented in Table 2. For the comparison of VSCI and VSOSI, the null hypothesis was rejected, which is to say that the ratios of sites deemed unacceptable to sites deemed acceptable were significantly different (p = 0.0005) for the two methods. For the comparison of the MAIS and VSOSI, the null hypothesis was not rejected, which is to say that the ratios of sites deemed unacceptable to sites deemed acceptable were not significantly different (p = 0.1473) for the two methods.

However, these chi-square tests only examined the numbers of sites classified each way, not the agreement and disagreement of classifications at the same individual sites. McNemar's test was used for the latter purpose (Zar 1999). The *null hypothesis* for

McNemar's test was that the proportion of individual study sites classified as having unacceptable ecological conditions by both the VSCI and VSOSI were not different (*i.e.,* the VSCI and VSOSI were in agreement for the designation of the same individual sites as unacceptable). In addition, the McNemar's test also reports the total percentage of sites in agreement (sites where both methods classified them as acceptable plus the sites where both methods classified them as unacceptable), for which the desirable outcome is the highest possible percentage agreement. Results of McNemar's test ($\propto = 0.05$) are presented in Table 3. For the comparison of VSCI and VSOSI, the null hypothesis was rejected, which is to say that the proportions of the same individual study sites classified as having unacceptable ecological conditions were significantly different ($p = 0.027$) for the two methods. There was only 52% agreement in the combined classification of streams (acceptable as well as unacceptable ecological condition designation) by the two methods. Note that in the 2 x 2 contingency tables shown in Table 1 (A) and (B), the top left and the bottom right cells show agreement of classification, unacceptable and acceptable, respectively. Hence for Table 1 (A), 9 + 2 = 11 sites out of 21, or 52% agreement. The top right and bottom left cells show the two types of disagreement in classification: one method classifies a site as acceptable while the other method classifies the site as unacceptable and vice versa. It would be preferable to have the two types of disagreement split equally, but in this case 9 of the 10 sites with disagreement were caused by the VSOSI classifying sites as acceptable while the VSCI classified those sites as impaired.

Similar to the chi-square analysis of ratios, the null hypothesis for McNemar's test was not rejected for the comparison of the MAIS and VSOSI, which is to say that the

proportions of individual study sites classified as having unacceptable ecological conditions were not significantly different ($p = 0.221$) for the two methods. There was 72% agreement in the combined classification of the streams by the two methods. It appeared that classifications of ecological condition made with the VSOSI tended to agree more closely with those made with the MAIS than the VSCI. However, the VSCI is the appropriate benchmark for comparison because that is the index that has been adopted by the VDEQ for biological monitoring of Virginia streams. The MAIS was included here only because it was the benchmark for the previous study by Engel and Voshell (2002). Therefore, further analyses were undertaken to understand the classification of streams by the volunteer VSOSI and the professional VSCI. The goal was to make the VSOSI classification agree more closely with the VSCI classification, without seriously weakening the agreement of VSOSI and MAIS classifications.

Our verifications of the taxonomic identifications made by the volunteers indicated that this was not the source of the problem. Seven of the 21 samples had no misidentifications or missed taxa. Six of the 21 samples had some taxa that were overlooked (were not identified or counted by the volunteers). In 5 of these 6 samples, only 1 individual was missed. In 1 of these 6 samples 2 individuals of the same taxon (Mayfly: Ephemeroptera) were missed. Most of the overlooked individuals were small, early instars. Taxa missed included 2 mayflies, 1 true bug, 1 gilled snail, 1 alderfly, 1 stonefly, and 1 true fly. In 7 of the 21 samples there were taxa counted and identified by volunteers that could not be located by the Virginia Tech investigators. "Other subsurface organisms" was the most common group that could not be located (5 out of the 7 samples). Other taxa that could not be located included "scuds," "most true flies,"

and “most caddisflies.” In 1 of the 21 samples there was a misidentification. About 50 beetle larvae (Hydrophilidae) were misidentified as “other subsurface organisms;” these should have been tallied as “beetles.” However, these mistakes were considered minor and would not have produced any substantive difference in the VSOSI calculation, certainly not a difference in the ecological condition classification.

It was speculated that the reason for the difference in ecological condition classification by the VSOSI might lie in the numerical cutoffs between unacceptable and acceptable. It can be seen in Table 1, for the VSOSI, five streams had a numerical score of 7, which is the minimum for acceptable classification, whereas a 6 would classify a stream as having unacceptable ecological condition. In addition, three streams had a numerical score of 8, which is only 2 numbers above the designation of unacceptable ecological condition. Since the disagreement between the volunteer and professional ecological condition designations consisted of too many sites being deemed acceptable by the volunteers, it was decided to analyze the effect of raising the VSOSI cutoff for acceptable to 8 and 9.

The ecological condition classification of the 21 study sites based upon the volunteer VSOSI cutoff being raised to 8 and 9 and the professional VSCI and MAIS remaining the same are presented in Tables 4 and 5, respectively. With the cutoff set at 8, the VSOSI classified 8 sites as unacceptable and 13 sites as acceptable; with a cutoff of 9, the VSOSI classified 11 sites as unacceptable and 10 sites as acceptable. This brought the ratio of unacceptable to acceptable sites in closer agreement with both the VSCI and the MAIS. The VSCI classified 11 sites as impaired and 10 sites as reference, which is exactly the same as the VSOSI with a cutoff of 9.

The ratios of numbers of sites classified as unacceptable versus acceptable by the adjusted cutoffs were analyzed with chi square goodness of fit tests, as had been done for the VSOSI with its original cutoff of 7. These results are presented in Tables 6 (cutoff = 8) and 7 (cutoff = 9). For both cutoffs, there were no significant differences in the observed frequency of unacceptable versus acceptable sites as determined by volunteers (VSOSI) and the expected frequency of unacceptable versus acceptable sites as determined by professional methods (VSCI, MAIS).

In addition, McNemar's test was used to test the proportion of the same individual study sites classified as having unacceptable ecological conditions by both the adjusted volunteer measures (VSOSI with cutoffs of 8 and 9) and the professional measures (VSCI, MAIS). These results are presented in Tables 8 and 9. The proportions of individual study sites classified as having unacceptable ecological conditions were not significantly different between the VSOSI with cutoffs of 8 or 9 and the two professional measures. The % agreement in the combined classification of the streams by the volunteer and professional measures was much higher with the modified cutoffs for the VSOSI. With a cutoff of 8, the VSOSI exhibited 76% agreement with the combined classification by the VSCI, and with a cutoff of 9 the VSOSI exhibited 81% agreement with the combined classification by the VSCI. The % agreement in the combined classification of the streams by the VSOSI and MAIS remained strong with the adjusted cutoffs. However, the agreement between the VSOSI and the MAIS in this study (71-76%) was not as strong as it was in the previous study (96%) by Engel and Voshell (2002).

Using a cutoff of 9 for the VSOSI, instead of the original cutoff of 7, would bring decisions about the ecological condition of streams made by volunteers using the VSOSI into very close agreement with the decisions made by professional biologists using the VSCI. However, in a recent draft document (Virginia Department of Environmental Quality 2006), VDEQ has proposed using confidence limits around the cutoff score of 60 for the VSCI to produce an area of uncertainty, or a "gray zone," in which there would be no decision about stream ecological condition. Streams achieving a score in this gray zone, currently proposed to be 55 to 63, would not be classified as either impaired or reference. This is a logical course of action, given the inherent variability in natural stream communities. Therefore, it seems prudent to explore a similar gray zone for the VSOSI in order to keep decisions about ecological condition of streams made with the two indices as comparable as possible. The complete VSOS database obtained since the modifications recommended by Engel and Voshell (2002) were implemented was analyzed by the same techniques that VDEQ (Virginia Department of Environmental Quality 2006) used to develop a gray zone based on the confidence limits around the cutoff. The complete database contained more than 2000 categorizations of stream sites by volunteer monitors using the VSOSI. Sites are different locations on streams, not necessarily different streams, but these locations were far enough apart that different ecological conditions might exist. However, some of these sites have been categorized many times by the same volunteers, while other sites have been categorized only once by different volunteers. Therefore, to avoid bias, the database was reduced to only one assessment per stream site, and it was decided *a priori* that this would be the first

assessment at a site. This created a database of 476 categorizations of ecological condition by volunteers using the VSOSI.

To initiate the analyses for a gray zone, a graph of the cumulative frequency distribution was plotted, including 95% confidence limits (Figure 2). Then, the 95% confidence interval and 95% confidence limits were extracted for each possible unit score of the VSOSI (Table 10). The possible unit scores for the VSOSI range from 1 to 12, but no stream had a score of 1 in the database. The VSOSI scores of most interest were the original cutoff of 7 and the proposed modified cutoffs of 8 and 9. It should first be noted that the 95% confidence intervals for VSOSI scores of 7, 8, and 9 were very small, less than 1 in each case (Table 10). Thus, there was no statistical justification for a gray zone that consists of more than 1 VSOSI unit score. The only question was whether the gray zone should be 7, 8, or 9.

The ecological condition classifications for the 21 validation sites based upon the VSOSI with gray zones of 7, 8, and 9 and the VSCI with a gray zone of 55 – 63 are presented in Tables 11, 12, and 13, respectively. The VSCI classified the 21 sites as follows: 6 unacceptable, 6 uncertain (gray zone), and 9 acceptable. Using a gray zone of 7, the VSOSI classified the 21 sites as follows: 3 unacceptable, 5 uncertain, and 13 acceptable. With a gray zone of 8, the VSOSI classified 8 sites as unacceptable, 3 as uncertain, and 10 as acceptable. With a gray zone of 9, the VSOSI classified 11 sites as unacceptable, 3 as uncertain, and 7 as acceptable.

The ratios of numbers of sites classified as unacceptable, uncertain, and acceptable by the VSOSI (observed) and VSCI (expected) were analyzed with chi square goodness of fit tests. These results are presented in Table 14 (A-C).). With the VSOSI

gray zone set at 9 (Table 14C), the observed ratio was significantly different from the expected ratio ($p = 0.0471$), which is the undesired outcome. With the VSOSI gray zone set at 7 and 8 (Table 14A and B, respectively), the observed ratio was not significantly different from the expected ratio ($p = 0.1787$ and $p = 0.3202$, respectively), both of which are the desired outcome. Of the two satisfactory choices for a gray zone, the VSOSI with a gray zone of 8 compares more favorably to the VSCI with a gray zone of 55=63. The *p* value for the chi square goodness of fit test is almost twice as high with the gray zone set at 8 compared to the gray zone set at 7. Most importantly, with the gray zone set at 7 the VSOSI categorized only 3 streams as unacceptable, which is half the number of streams categorized as unacceptable by the VSCI. In contrast, with the gray zone set at 8 the VSOSI categorized 8 streams as unacceptable, which is slightly more than the number of streams categorized as unacceptable by the VSCI.

Summary and Recommendations

The modified SOS protocol that was developed by Engel and Voshell (2002) does not need to be changed in regard to sampling, identification, enumeration, and calculation of the multimetric index called the Virginia Save-Our-Streams Index (VSOSI). The volunteers made very few mistakes in the identification of macroinvertebrates. These mistakes were considered minor and would not produce any substantive difference in the VSOSI calculation, certainly not a difference in the ecological condition classification. The VSOSI correlates very strongly with the Virginia Stream Condition Index (VSCI) used by professional biologists at the Virginia Department of Environmental Quality. However, in this validation study the VSOSI did not agree satisfactorily with the classification of stream ecological condition done by professional biologists using the

VSCI. The VSOSI overrated too many streams (*i.e.*, classified them as acceptable, when the VSCI classified them as impaired). A simple solution to this situation was found: raise the numerical value required for the VSOSI to classify a stream as acceptable. Using a cutoff of 9 for a stream to be classified as acceptable by the VSOSI agreed very closely (81%) with the VSCI classification of the same streams. In addition, the disagreement of site classification was equally split between classifying reference as unacceptable and classifying impaired as acceptable.

We recommend that Virginia Save-Our-Streams continue to use the existing protocol as modified by Engel and Voshell (2002) and to calculate the same VSOSI, but to shift the criterion for acceptable ecological condition to 9. If an uncertain ("gray zone") is desired for the VSOSI to be comparable to recently suggested modifications of the VSCI, then we recommend that the gray zone be the VSOSI unit score of 8. Classification of the ecological condition of streams by the VSOSI with a gray zone of 8 agreed very well with the VSCI with a gray zone of 55-63. With or without the gray zone, all data collected since the modification by Engel and Voshell (2002) are still valid. The ecological condition classifications merely need to be reassigned based on an acceptable cutoff of 9 and possibly a gray zone of 8 in order to be in agreement with the VSCI classification of reference conditions.

The results of the current validation study are not unexpected because the VSCI was not available at the time of the previous study. Lastly, we recommend that Virginia Save-Our-Streams periodically revalidate the performance of the VSOSI against the VSCI because VDEQ is still analyzing and validating the VSCI.

References

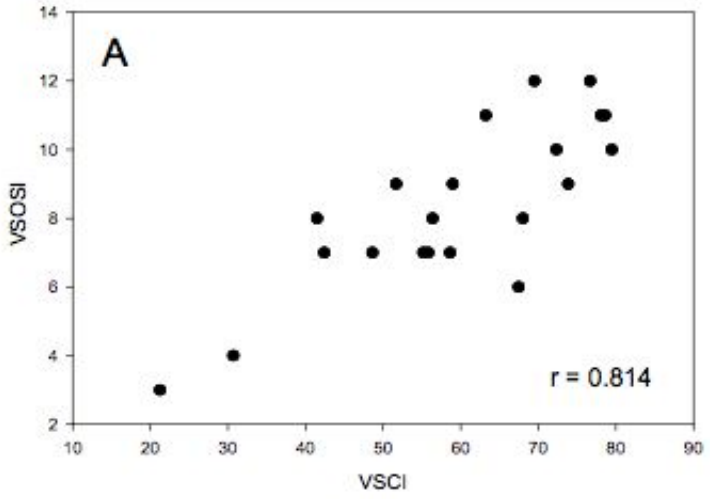
Burton, J. and J. Gerritsen. 2003. A stream condition index for Virginia non-coastal streams. Report prepared by Tetra Tech, Inc. for USEPA Office of Science and Technology, Office of Water, Washington, DC; USEPA Region 3 Environmental Services Division, Wheeling, WV; Virginia Department of Environmental Quality, Richmond, VA.

Engel, S. R. and J. R. Voshell, Jr. 2002. Volunteer biological monitoring: can it accurately assess the ecological condition of streams? American Entomologist 48: 164-177.

Virginia Department of Environmental Quality. 2006. Using probabilistic monitoring data to validate the non-coastal Virginia Stream Condition Index. VDEQ Technical Bulletin WQA/2006-001. Water Quality Monitoring, Biological Monitoring and Water Quality Assessment Programs, Department of Environmental Quality, Richmond, VA. (Draft)

Zar, J. H. 1999. Biostatistical analysis, 4th ed. Prentice Hall, Upper Saddle River, New Jersey. 663 pp.

Figure 1. Results of Pearson product-moment correlation analyses. (A) Compares Virginia Save-Our-Streams Index (VSOSI) used by volunteers to Virginia Stream Condition Index (VSCI) used by professional biologists at the Virginia Department of Environmental Quality. (B) Compares Virginia Save-Our-Streams Index (VSOSI) to the Macroinvertebrate Aggregated Index for Streams (MAIS) developed for use by professional biologists.



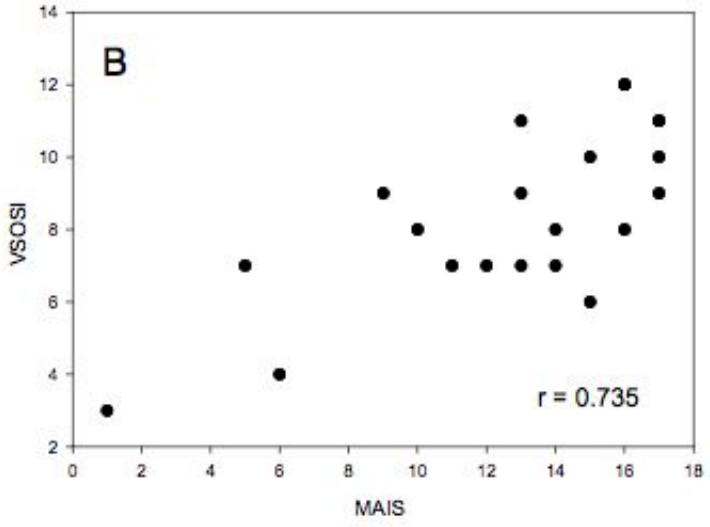


Figure 2. Graph of the cumulative frequency distribution for assessments made by volunteers at 476 stream sites using the Virginia Save-Our-Streams Index. Only one assessment per site (the first one) was included in the analysis.

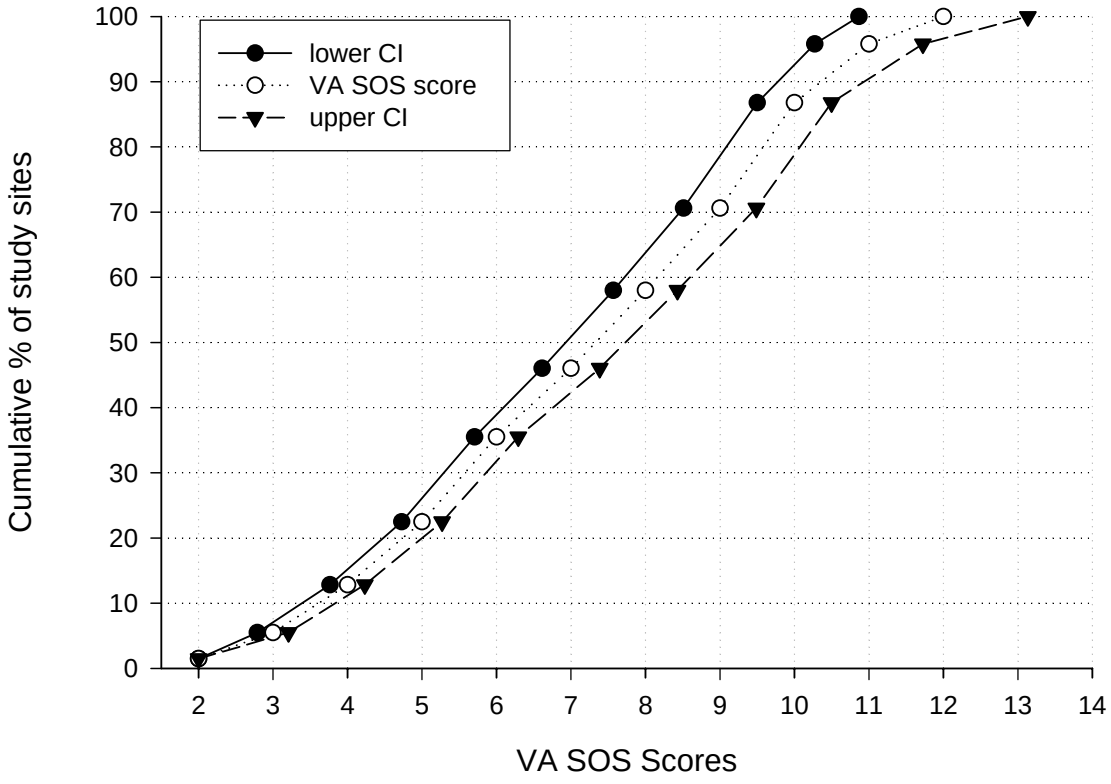


Table 1. List of all study sites and the scores and resulting ecological condition category for the multimetric indices that were compared in this study. VSOSI = Virginia Save-Our-Streams Index (volunteers); VSCI = Virginia Stream Condition Index (used by professional biologists at the Virginia Department of Environmental Quality); MAIS = Macroinvertebrate Aggregated Index for Streams (developed for use by professional biologists). The cutoffs for ecological condition category were: **VSOSI ≥7 = Acceptable**); VSCI ≥60 = Reference; MAIS ≥13 = Acceptable.

	VSOSI		**VSCI**		**MAIS**	
Site	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category
Black's Run	7	Acceptable	42	Impaired	5	Unacceptable
Buffalo Creek	7	Acceptable	55	Impaired	11	Unacceptable
Cedar Creek	7	Acceptable	59	Impaired	14	Acceptable
Cook's Creek	3	Unacceptable	21	Impaired	1	Unacceptable
Crab Creek	8	Acceptable	41	Impaired	10	Unacceptable
Craig Creek	9	Acceptable	74	Reference	17	Acceptable
Fiery Run	6	Unacceptable	67	Reference	15	Acceptable
unnamed trib to Goose Creek	12	Acceptable	77	Reference	16	Acceptable
Lewis Creek	9	Acceptable	59	Impaired	13	Acceptable
Lick Run	11	Acceptable	79	Reference	17	Acceptable
Little Creek	8	Acceptable	56	Impaired	14	Acceptable
Middle River	4	Unacceptable	31	Impaired	6	Unacceptable
Passage Creek	11	Acceptable	63	Reference	13	Acceptable
Peak Creek	9	Acceptable	52	Impaired	9	Unacceptable
Roanoke River	7	Acceptable	56	Impaired	13	Acceptable
S.F. Holston River	12	Acceptable	69	Reference	16	Acceptable
S.F. Little Chestnut Creek	11	Acceptable	78	Reference	17	Acceptable
S.F. Roanoke River	8	Acceptable	68	Reference	16	Acceptable
Teels Creek	7	Acceptable	49	Impaired	12	Unacceptable
Toms Creek	10	Acceptable	79	Reference	17	Acceptable
Wolf Creek	10	Acceptable	72	Reference	15	Acceptable

Table 2. Chi square goodness of fit test ($\alpha = 0.05$) comparing number of unacceptable versus acceptable determinations of ecological condition based on the VSOSI with a cutoff of 7.

(A) VSOSI (≥7 = Acceptable) versus VSCI (≥60 = Reference).

Ecological Condition Categories	VSOSI (Observed Frequency)	VSCI (Expected Frequency)
Unacceptable	3	11
Acceptable	18	10
χ^2 value = 12.218		
p value = 0.0005		

(B) VSOSI (≥7 = Acceptable) versus MAIS (≥13 = Acceptable).

Ecological Condition Categories	VSOSI (Observed Frequency)	MAIS (Expected Frequency)
Unacceptable	3	7
Acceptable	18	14
χ^2 value = 2.100		
p value = 0.1473		

Table 3. Classification analysis with the McNemar test ($\alpha = 0.05$) comparing conclusions about ecological condition based on the VSOSI with a cutoff of 7.

(A) VSOSI (≥7 = Acceptable) versus VSCI (≥60 = Reference)

	VSOSI		
	Unacceptable		Acceptable
VSCI			
Impaired	2		1
Reference	9		9
% agreement		52 %	
McNemar test *p*-value		0.027	

(B) VSOSI (≥7 = Acceptable) versus MAIS (≥13 = Acceptable).

	VSOSI		
	Unacceptable		Acceptable
MAIS			
Unacceptable	13		1
Acceptable	5		2
% agreement		72 %	
McNemar test *p*-value		0.221	

Table 4. List of all study sites and the scores and resulting ecological condition category for the multimetric indices that were compared in this study. VSOSI = Virginia Save-Our-Streams Index (volunteers); VSCI = Virginia Stream Condition Index (used by professional biologists at the Virginia Department of Environmental Quality); MAIS = Macroinvertebrate Aggregated Index for Streams (developed for use by professional biologists). The cutoffs for ecological condition category were: **VSOSI ≥8 = Acceptable**); VSCI ≥60 = Reference; MAIS ≥13 = Acceptable.

	VSOSI		**VSCI**		**MAIS**	
Site	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category
Black's Run	7	Unacceptable	42	Impaired	5	Unacceptable
Buffalo Creek	7	Unacceptable	55	Impaired	11	Unacceptable
Cedar Creek	7	Unacceptable	59	Impaired	14	Acceptable
Cook's Creek	3	Unacceptable	21	Impaired	1	Unacceptable
Crab Creek	8	Acceptable	41	Impaired	10	Unacceptable
Craig Creek	9	Acceptable	74	Reference	17	Acceptable
Fiery Run	6	Unacceptable	67	Reference	15	Acceptable
unnamed trib to Goose Creek	12	Acceptable	77	Reference	16	Acceptable
Lewis Creek	9	Acceptable	59	Impaired	13	Acceptable
Lick Run	11	Acceptable	79	Reference	17	Acceptable
Little Creek	8	Acceptable	56	Impaired	14	Acceptable
Middle River	4	Unacceptable	31	Impaired	6	Unacceptable
Passage Creek	11	Acceptable	63	Reference	13	Acceptable
Peak Creek	9	Acceptable	52	Impaired	9	Unacceptable
Roanoke River	7	Unacceptable	56	Impaired	13	Acceptable
S.F. Holston River	12	Acceptable	69	Reference	16	Acceptable
S.F. Little Chestnut Creek	11	Acceptable	78	Reference	17	Acceptable
S.F. Roanoke River	8	Acceptable	68	Reference	16	Acceptable
Teels Creek	7	Unacceptable	49	Impaired	12	Unacceptable
Toms Creek	10	Acceptable	79	Reference	17	Acceptable
Wolf Creek	10	Acceptable	72	Reference	15	Acceptable

Table 5. List of all study sites and the scores and resulting ecological condition category for the multimetric indices that were compared in this study. VSOSI = Virginia Save-Our-Streams Index (volunteers); VSCI = Virginia Stream Condition Index (used by professional biologists at the Virginia Department of Environmental Quality); MAIS = Macroinvertebrate Aggregated Index for Streams (developed for use by professional biologists). The cutoffs for ecological condition category were: **VSOSI ≥9 = Acceptable**); VSCI ≥60 = Reference; MAIS ≥13 = Acceptable.

	VSOSI		**VSCI**		**MAIS**	
Site	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category
Black's Run	7	Unacceptable	42	Impaired	5	Unacceptable
Buffalo Creek	7	Unacceptable	55	Impaired	11	Unacceptable
Cedar Creek	7	Unacceptable	59	Impaired	14	Acceptable
Cook's Creek	3	Unacceptable	21	Impaired	1	Unacceptable
Crab Creek	8	Unacceptable	41	Impaired	10	Unacceptable
Craig Creek	9	Acceptable	74	Reference	17	Acceptable
Fiery Run	6	Unacceptable	67	Reference	15	Acceptable
unnamed trib to Goose Creek	12	Acceptable	77	Reference	16	Acceptable
Lewis Creek	9	Acceptable	59	Impaired	13	Acceptable
Lick Run	11	Acceptable	79	Reference	17	Acceptable
Little Creek	8	Unacceptable	56	Impaired	14	Acceptable
Middle River	4	Unacceptable	31	Impaired	6	Unacceptable
Passage Creek	11	Acceptable	63	Reference	13	Acceptable
Peak Creek	9	Acceptable	52	Impaired	9	Unacceptable
Roanoke River	7	Unacceptable	56	Impaired	13	Acceptable
S.F. Holston River	12	Acceptable	69	Reference	16	Acceptable
S.F. Little Chestnut Creek	11	Acceptable	78	Reference	17	Acceptable
S.F. Roanoke River	8	Unacceptable	68	Reference	16	Acceptable
Teels Creek	7	Unacceptable	49	Impaired	12	Unacceptable
Toms Creek	10	Acceptable	79	Reference	17	Acceptable
Wolf Creek	10	Acceptable	72	Reference	15	Acceptable

Table 6. Chi square goodness of fit test ($\alpha = 0.05$) comparing number of unacceptable versus acceptable determinations of ecological condition based on the VSOSI with a cutoff of 8

(A) VSOSI (≥8 = Acceptable) versus VSCI (≥60 = Reference).

Ecological Condition Categories	VSOSI (Observed Frequency)	VSCI (Expected Frequency)
Unacceptable	8	11
Acceptable	13	10
χ^2 value = 1.718		
p value = 0.1899		

(B) VSOSI (≥8 = Acceptable) versus MAIS (≥13 = Acceptable).

Ecological Condition Categories	VSOSI (Observed Frequency)	MAIS (Expected Frequency)
Unacceptable	8	7
Acceptable	13	14
χ^2 value = 0.104		
p value = 0.7471		

Table 7. Chi square goodness of fit test ($\alpha = 0.05$) comparing number of unacceptable versus acceptable determinations of ecological condition based on the VSOSI with a cutoff of 9

(A) VSOSI (≥9 = Acceptable) versus VSCI (≥60 = Reference).

Ecological Condition Categories	VSOSI (Observed Frequency)	VSCI (Expected Frequency)
Unacceptable	11	11
Acceptable	10	10
χ^2 value = 0.000		
p value = 1.0000		

(B) VSOSI (≥9 = Acceptable) versus MAIS (≥13 = Acceptable).

Ecological Condition Categories	VSOSI (Observed Frequency)	MAIS (Expected Frequency)
Unacceptable	11	7
Acceptable	10	14
χ^2 value = 1.556		
p value = 0.2123		

Table 8. Classification analysis with the McNemar test ($\alpha = 0.05$) comparing conclusions about ecological condition based on the **VSOSI with a cutoff of 8**.

(A) VSOSI (≥8 = Acceptable) versus VSCI (≥60 = Reference).

	VSOSI		
	Unacceptable		Acceptable
Virginia SCI			
Impaired	7		1
Reference	4		9
% agreement		76 %	
McNemar test *p*-value		0.371	

(B) VSOSI (≥8 = Acceptable) versus MAIS (≥13 = Acceptable).

	VSOSI		
	Unacceptable		Acceptable
MAIS			
Unacceptable	5		2
Acceptable	3		11
% agreement		76 %	
McNemar test *p*-value		1.00	

Table 9. Classification analysis with the McNemar test ($\alpha = 0.05$) comparing conclusions about ecological condition based on the **VSOSI with a cutoff of 9**.

(A) VSOSI (≥9 = Acceptable) versus VSCI (≥60 = Reference).

	VSOSI		
	Unacceptable		Acceptable
Virginia SCI			
Impaired	9		2
Reference	2		8
% agreement		81 %	
McNemar test *p*-value		0.617	

(B) VSOSI (≥9 = Acceptable) versus MAIS (≥13 = Acceptable).

	VSOSI		
	Unacceptable		Acceptable
MAIS			
Unacceptable	6		1
Acceptable	5		9
% agreement		71 %	
McNemar test *p*-value		0.221	

Table 10. 95% confidence intervals and 95% confidence limits for each possible unit score for the Virginia Save-Our-Streams Index (VSOSI) based on assessments made by volunteers at 476 stream sites. Only one assessment per site (the first one) was included in the analysis.

Number of Sites (Cumulative %)	VSOSI Score	Confidence Interval	Lower Confidence Limit	Upper Confidence Limit
1	2	±0.00	2.00	2.00
5	3	±0.21	2.79	3.21
13	4	±0.23	3.77	4.23
22	5	±0.27	4.73	5.27
36	6	±0.29	5.71	6.29
46	7	±0.39	6.61	7.39
58	8	±0.43	7.57	8.43
71	9	±0.49	8.51	9.49
87	10	±0.50	9.50	10.50
96	11	±0.70	10.27	11.70
100	12	±1.13	10.87	13.13

Table 11. List of all study sites and the scores and resulting ecological condition category for the two major multimetric indices that were compared in this study, with an intermediate uncertain ecological condition category (gray zone). VSOSI = Virginia Save-Our-Streams Index (volunteers); VSCI = Virginia Stream Condition Index (used by professional biologists at the Virginia Department of Environmental Quality). The cutoffs for ecological condition category were: **VSOSI 1-6 = Unacceptable, 7 = Uncertain (Gray Zone), ≥8 = Acceptable); VSCI 1-54 = Impaired, 55-63 = Uncertain (Gray Zone), ≥64 = Reference**.

	VSOSI		**VSCI**	
Site	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category
Black's Run	7	Uncertain (Gray Zone)	42	Impaired
Buffalo Creek	7	Uncertain (Gray Zone)	55	Uncertain (Gray Zone)
Cedar Creek	7	Uncertain (Gray Zone)	59	Uncertain (Gray Zone)
Cook's Creek	3	Unacceptable	21	Impaired
Crab Creek	8	Acceptable	41	Impaired
Craig Creek	9	Acceptable	74	Reference
Fiery Run	6	Unacceptable	67	Reference
unnamed trib to Goose Creek	12	Acceptable	77	Reference
Lewis Creek	9	Acceptable	59	Uncertain (Gray Zone)
Lick Run	11	Acceptable	79	Reference
Little Creek	8	Acceptable	56	Uncertain (Gray Zone)
Middle River	4	Unacceptable	31	Impaired
Passage Creek	11	Acceptable	63	Uncertain (Gray Zone)
Peak Creek	9	Acceptable	52	Impaired
Roanoke River	7	Uncertain (Gray Zone)	56	Uncertain (Gray Zone)
S.F. Holston River	12	Acceptable	69	Reference
S.F. Little Chestnut Creek	11	Acceptable	78	Reference
S.F. Roanoke River	8	Acceptable	68	Reference
Teels Creek	7	Uncertain (Gray Zone)	49	Impaired
Toms Creek	10	Acceptable	79	Reference
Wolf Creek	10	Acceptable	72	Reference

Table 12. List of all study sites and the scores and resulting ecological condition category for the two major multimetric indices that were compared in this study, an intermediate uncertain ecological condition category (gray zone). VSOSI = Virginia Save-Our-Streams Index (volunteers); VSCI = Virginia Stream Condition Index (used by professional biologists at the Virginia Department of Environmental Quality). The cutoffs for ecological condition category were: **VSOSI 1-7 = Unacceptable, 8 = Uncertain (Gray Zone), ≥9 = Acceptable); VSCI 1-54 = Impaired, 55-63 = Uncertain (Gray Zone), ≥64 = Reference**.

	VSOSI		**VSCI**	
Site	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category
Black's Run	7	Unacceptable	42	Impaired
Buffalo Creek	7	Unacceptable	55	Uncertain (Gray Zone)
Cedar Creek	7	Unacceptable	59	Uncertain (Gray Zone)
Cook's Creek	3	Unacceptable	21	Impaired
Crab Creek	8	Uncertain (Gray Zone)	41	Impaired
Craig Creek	9	Acceptable	74	Reference
Fiery Run	6	Unacceptable	67	Reference
unnamed trib to Goose Creek	12	Acceptable	77	Reference
Lewis Creek	9	Acceptable	59	Uncertain (Gray Zone)
Lick Run	11	Acceptable	79	Reference
Little Creek	8	Uncertain (Gray Zone)	56	Uncertain (Gray Zone)
Middle River	4	Unacceptable	31	Impaired
Passage Creek	11	Acceptable	63	Uncertain (Gray Zone)
Peak Creek	9	Acceptable	52	Impaired
Roanoke River	7	Unacceptable	56	Uncertain (Gray Zone)
S.F. Holston River	12	Acceptable	69	Reference
S.F. Little Chestnut Creek	11	Acceptable	78	Reference
S.F. Roanoke River	8	Uncertain (Gray Zone)	68	Reference
Teels Creek	7	Unacceptable	49	Impaired
Toms Creek	10	Acceptable	79	Reference
Wolf Creek	10	Acceptable	72	Reference

Table 13. List of all study sites and the scores and resulting ecological condition category for the two major multimetric indices that were compared in this study, with an intermediate uncertain ecological condition category (gray zone). VSOSI = Virginia Save-Our-Streams Index (volunteers); VSCI = Virginia Stream Condition Index (used by professional biologists at the Virginia Department of Environmental Quality). The cutoffs for ecological condition category were: **VSOSI 1-8 = Unacceptable, 9 = Uncertain (Gray Zone), ≥10 = Acceptable); VSCI 1-54 = Impaired, 55-63 = Uncertain (Gray Zone), ≥64 = Reference**.

	VSOSI		VSCI	
Site	Numerical Score	Ecological Condition Category	Numerical Score	Ecological Condition Category
Black's Run	7	Unacceptable	42	Impaired
Buffalo Creek	7	Unacceptable	55	Uncertain (Gray Zone)
Cedar Creek	7	Unacceptable	59	Uncertain (Gray Zone)
Cook's Creek	3	Unacceptable	21	Impaired
Crab Creek	8	Unacceptable	41	Impaired
Craig Creek	9	Uncertain (Gray Zone)	74	Reference
Fiery Run	6	Unacceptable	67	Reference
unnamed trib to Goose Creek	12	Acceptable	77	Reference
Lewis Creek	9	Uncertain (Gray Zone)	59	Uncertain (Gray Zone)
Lick Run	11	Acceptable	79	Reference
Little Creek	8	Unacceptable	56	Uncertain (Gray Zone)
Middle River	4	Unacceptable	31	Impaired
Passage Creek	11	Acceptable	63	Uncertain (Gray Zone)
Peak Creek	9	Uncertain (Gray Zone)	52	Impaired
Roanoke River	7	Unacceptable	56	Uncertain (Gray Zone)
S.F. Holston River	12	Acceptable	69	Reference
S.F. Little Chestnut Creek	11	Acceptable	78	Reference
S.F. Roanoke River	8	Unacceptable	68	Reference
Teels Creek	7	Unacceptable	49	Impaired
Toms Creek	10	Acceptable	79	Reference
Wolf Creek	10	Acceptable	72	Reference

Table 14. Chi square goodness of fit test ($\alpha = 0.05$) comparing number of unacceptable, uncertain (gray zone), and acceptable determinations of ecological condition based on the VSOSI versus the VSCI. In each case the cutoffs for ecological condition with the VSCI were: 55-63 = Uncertain (Gray Zone), ≤54 = Impaired, ≥64 = Reference.

A) Cutoffs for ecological condition category with VSOSI were: 7 = Uncertain (Gray Zone), ≤6 = Unacceptable, ≥8 = Acceptable.

Ecological Condition Categories	VSOSI (Observed Frequency)	VSCI (Expected Frequency)
Unacceptable	3	6
Uncertain (Gray Zone)	5	6
Acceptable	13	9
χ^2 value = 3.444		
p value = 0.1787		

B) Cutoffs for ecological condition category with VSOSI were: 8 = Uncertain (Gray Zone), ≤7 = Unacceptable, ≥9 = Acceptable.

Ecological Condition Categories	VSOSI (Observed Frequency)	VSCI (Expected Frequency)
Unacceptable	8	6
Uncertain (Gray Zone)	3	6
Acceptable	10	9
χ^2 value = 2.278		
p value = 0.3202		

C) Cutoffs for ecological condition category with VSOSI were: 9 = Uncertain (Gray Zone), ≤8 = Unacceptable, ≥10 = Acceptable.

Ecological Condition Categories	VSOSI (Observed Frequency)	VSCI (Expected Frequency)
Unacceptable	11	6
Uncertain (Gray Zone)	3	6
Acceptable	7	9
χ^2 value = 6.111		
p value = 0.0471		