An index of biotic integrity for macroinvertebrate stream bioassessment conducted by community scientists

Patrick M. Edwards[1,7], **Daniel Bedell**[1,8], **Shannon L. Hubler**[2,9], **Chad A. Larson**[3,10], **Kate H. Macneale**[4,11], **Elisa Mickelson**[1,12], **Chris Prescott**[5,13], **Elinore Webb**[6,14], **and Jo Wilhelm**[4,15]

[1]Department of Environmental Science and Management, Portland State University, Portland, Oregon, USA
[2]Oregon Department of Environmental Quality, Hillsboro, Oregon, USA
[3]Washington State Department of Ecology, Environmental Assessment Program, Olympia, Washington, USA
[4]King County Department of Natural Resources and Parks, Seattle, Washington, USA
[5]City of Portland Bureau of Environmental Services, Portland, Oregon, USA
[6]Geography Department, Portland State University, Portland, Oregon, USA

Abstract: Community science bioassessment has great potential to inform comprehensive stream management plans, but regional analytical tools are needed to evaluate macroinvertebrate data collected through community science programs. To this end, we modified a pre-existing professional index of biotic integrity (IBI) to create a community science IBI (CS-IBI), designed for stream macroinvertebrate data collected by community scientists with minimal training. We used data collected by both professional and community scientists to develop, calibrate, and validate the CS-IBI at 76 stream sites in the Puget Lowland and Willamette Valley ecoregions of the Pacific Northwest in the United States. Community science data were taxonomically coarser and more variable than data generated by professionals; however, IBI scores and assemblage data were statistically similar between community science and professional data. Stream impairment categories classified by family-level CS-IBI scores matched genus-level professional classifications 65% of the time and never diverged by >1 category. CS-IBI scores were negatively related to the percentage of agriculture and land development in the watershed, although this relationship was weaker than for professional IBI scores. Despite increased variability in data generated by community scientists, our findings suggest the CS-IBI performs similarly to a professional IBI across a gradient of human influence. Although we do not advocate using the CS-IBI in regulatory settings, we believe the development of community science IBIs enhances, expands, and strengthens public partnerships, thereby supporting environmental managers' efforts to monitor and restore degraded streams and rapidly respond to pollution events. Our hope is that the CS-IBI will improve the applicability of community science bioassessment data and serve as a model for how agencies can develop regionalized macroinvertebrate IBIs for use in comprehensive watershed management plans.

Key words: citizen science, community science, stream macroinvertebrates, stream bioassessment, index of biotic integrity, watershed stressors

Community science (also referred to as citizen science) is increasingly being incorporated into the environmental and ecological sciences (Fraisl et al. 2022). There is a wide range of definitions for contributory environmental community science, but here we consider it to be a partnership between community scientists and professional scientists with the aim of generating data for scientific investigation and environmental management while also engaging the public in the scientific process (Bonney et al. 2009, Silvertown 2009, Thornhill et al. 2019). Aside from generating data, there are other compelling reasons for establishing environmentally focused community science programs, including natural resource education (Bonney 2021), raising awareness of publicly funded projects to improve surface water quality (Bonney et al. 2016, Ballard et al. 2017, Walker et al. 2021), engaging the public in environmental policy and management

E-mail addresses: [7]patrick.edwards@pdx.edu; [8]dbedell@pdx.edu; [9]shannon.l.hubler@state.or.us; [10]clar461@ecy.wa.gov; [11]Kate.Macneale@kingcounty.gov; [12]emick2@pdx.edu; [13]Chris.Prescott@portlandoregon.gov; [14]elinore@pdx.edu; [15]Jo.Wilhelm@kingcounty.gov

Received 31 August 2022; Accepted 13 December 2023; Published online 25 April 2024. Associate Editor, Daren Carlisle.

decisions (Overdevest et al. 2004, McKinley et al. 2017, Edwards and Shaloum 2020, Burdette et al. 2021), and providing authentic research experiences that increase participation and retention of underrepresented students in science, technology, engineering, and math education (Pandya 2012, NASEM 2018, Valle et al. 2021).

Community science has long played a role in monitoring the condition of freshwater systems (Firehock and West 1995), and there is growing interest in incorporating community science bioassessments in comprehensive stream and watershed management (Rieman et al. 2015, Hopfensperger et al. 2021, White et al. 2021). One way community science can be incorporated is via macroinvertebrate bioassessment, which is the systematic evaluation of macroinvertebrate assemblages to assess stream ecological condition (Rosenberg and Resh 1993). Macroinvertebrates are the most widely used taxonomic group to assess the ecological condition of streams and rivers (Resh 2008). Macroinvertebrates are also ideal for data collection by nonscientists because they are easy to collect and interesting to observe. A community science approach to macroinvertebrate bioassessment provides participants with authentic research experiences through the use of simple and inexpensive field sampling techniques while generating biological datasets that can be used to assess stream condition in real time. Moreover, macroinvertebrates are more engaging than physical and chemical measures, and macroinvertebrate bioassessment is an effective approach for natural science education (Bonney et al. 2009). Although many resources have been developed for community science macroinvertebrate bioassessment (Firehock 1994, Walk 1997, Edwards 2016, White et al. 2021), there remains a need for regional tools to use the data collected by community scientists to evaluate stream condition within a framework that is useful to stream managers (Callaghan et al. 2019).

One such tool is an index of biotic integrity (IBI). IBIs are analytical tools used as indicators of environmental conditions. They include multiple biological metrics characterizing both the taxonomic and functional aspects of the biotic community (Karr 1998, Stoddard et al. 2008). Macroinvertebrate IBIs are frequently used in stream bioassessment and are often used for regulatory purposes (Karr 1998, Davies and Jackson 2006). IBIs are useful for community science because they are simple to calculate, easy to interpret, and can be used as an educational tool for increasing environmental stewardship and knowledge of stream conditions. However, IBIs are generally applied to data collected by trained scientists, and there are challenges in using IBIs in community science related to how community scientists collect, process, and enumerate macroinvertebrate samples. Community science groups often sort, identify, and count live macroinvertebrates in the field under ambient light conditions with little to no magnification (Nerbonne et al. 2008, Edwards 2016). In contrast, professional scientists sort, identify, and enumerate preserved macroinvertebrates in the laboratory under magnification and artificial lighting (Nerbonne et al. 2008). As a result, live sorting and counting by community scientists generates macroinvertebrate data that are biased towards large or motile macroinvertebrates that are easy to see in the field (Nerbonne et al. 2008). For example, community scientists detect fewer taxa overall and fewer small organisms than professionals (Edwards 2016). Consequently, IBIs designed for professionally collected biological data are unreliable when calculated using macroinvertebrate data collected by community scientists (O'Leary et al. 2004, Nerbonne et al. 2008).

Few macroinvertebrate metrics or indexes have been developed for community science bioassessment. Most of these were developed using a lethal field method or were based on data collected by professional scientists, were not validated with community science data, and have not been formally published (e.g., Water Assessment by Volunteer Evaluators, Wai Care Invertebrate Monitoring Protocol; Aukland Council 2013, Onion et al. 2023). The only peer-reviewed paper we are aware of that describes a metric for community science bioassessment developed using data collected by community scientists was published by Pinto et al. (2020), who developed a pollution tolerance-based macroinvertebrate metric for community science bioassessment that correlated with the ecological status of Portuguese streams and corresponded with data collected by professionals. However, from the perspective of a stream manager, a common limitation of all the published and unpublished studies we reviewed was a lack of independent validation of the metric and estimation of metric variability. These considerations are important when evaluating the degree of confidence stream managers can expect in macroinvertebrate data generated by community scientists (Brown and Williams 2019). Furthermore, the metrics developed for community science were all based on the pollution tolerance of macroinvertebrates and did not include other biological traits, such as feeding, which are critical to understanding the biological integrity of streams (Karr and Chu 2000, Davies and Jackson 2006). Moreover, none of the community science metrics were designed to correspond with a professional IBI, limiting their value for environmental management.

To address these limitations, we developed a regionally specific community science IBI (CS-IBI). The CS-IBI is an analytical tool specifically designed for macroinvertebrate bioassessment data collected by community scientists sampling streams in the Puget Lowland and Willamette Valley ecoregions. In this paper, we describe the development, calibration, and validation of the CS-IBI, predicting that it would perform similarly to an IBI used by professional scientists for regulatory purposes. The main objectives of our study were to 1) develop and calibrate the CS-IBI across a gradient of watershed disturbance in 2 ecoregions in the western United States, 2) validate the CS-IBI by comparing it with a professionally derived IBI using data collected simultaneously by professional scientists from the same streams, 3) evaluate the reliability of the CS-IBI data generated by community scientists by assessing its variability over time, and 4) compare macroinvertebrate assemblage data generated by a nonlethal sampling method designed for community science with assemblage data generated by standard macroinvertebrate sampling methods. We also aimed to identify key issues related to the use of community science bioassessment data in comprehensive stream management plans.

METHODS

To develop, calibrate, and validate a regionally specific CS-IBI, we conducted a 3-part study to generate 4 datasets (Table 1) of macroinvertebrate samples from streams across the Puget Lowland and Willamette Valley ecoregions of the Pacific Northwest (PNW), United States (Fig. 1). In the 1st part, we developed and calibrated the CS-IBI with data derived from macroinvertebrates nonlethally sampled and identified by professional scientists. In the 2nd part, we validated the CS-IBI with a separate dataset of macroinvertebrates sampled and identified by professional scientists. In this step, scientists sampled paired riffles—1 riffle with lethal methods and the other with nonlethal methods developed for community scientists—then scored them with a professional IBI and the CS-IBI, respectively, allowing comparison of the scores between the 2 indices and community composition between the 2 datasets. Finally, we used a long-term macroinvertebrate dataset collected by community scientists, in this case students, to test the reliability of the CS-IBI over multiple sampling years. We then used linear regression to model the relationships between the IBI scores from all 4 datasets with watershed disturbance at each site.

For the purposes of this study, we define a community scientist as a student, volunteer, or other member of the public who is not professionally or academically trained as a scientist. We define a professional scientist as a trained scientist with an academic or professional affiliation. We recognize that many community scientists are also retired academics, environmental scientists, and natural resource managers, but no participants involved in this study fit these categories.

Study area

The Puget Lowland and Willamette Valley ecoregions represent the largest population centers of the PNW and contain critical habitat for endangered salmonids (Naiman and Bilby 1998). In general, these regions have mild wet winters and cool dry summers. Historically, streams in these regions were generally cold with gravel and cobble substrates used by salmon for spawning; however, urbanization, agriculture, and other stressors such as logging and mining have degraded many of these stream ecosystems (Wilson and Sorenson 2012). The development of a regional CS-IBI for the Puget Lowland and the Willamette Valley is important because macroinvertebrate communities are highly variable across the landscape and, thus, require regional adjustments across diagnostic indices (Miller et al. 1988). Furthermore, the Puget Lowland and Willamette Valley ecoregions have >90 watershed councils tasked with monitoring streams (PME, unpublished data), constituting the highest density of active community science programs in the PNW region.

We selected streams across a gradient of watershed conditions from a pool of stream sites routinely sampled by state or local agencies. We used the 2016 National Landcover Database (Dewitz 2019) to characterize land use by calculating the percentage of the watershed area upstream of the sample site that was predominantly agricultural (%hay or pasture and %cultivated crops) or developed (%low, %medium, and %high development). We summed these values to determine the total amount of agriculture and developed land use (%Ag+Dev) in the watershed upstream of each site.

Table 1. Summary and description of the datasets used for this study. B-IBI = benthic index of biotic integrity, CS-IBI = community science index of biotic integrity.

Dataset	Collectors	Purpose	Description	*N*	Dates collected
CS-IBI calibration	Professional scientists	Calibration	Collected using a nonlethal field method. Family-level data used to develop and calibrate the CS-IBI.	48 streams	Autumn 2020
B-IBI validation	Professional scientists	Validation	Collected using a standard professional method. Genus-level data compared with CS-IBI validation scores.	20 streams	Autumn 2019
CS-IBI validation	Professional scientists	Validation	Collected using a nonlethal field method simultaneously with B-IBI samples. Family-level data compared with B-IBI validation scores.	20 streams	Autumn 2019
Student	Middle school, high school, or university students	Validation	Collected using a nonlethal field method. Family-level data used to evaluate CS-IBI scores.	10 streams, 90 sampling events	2005–2021

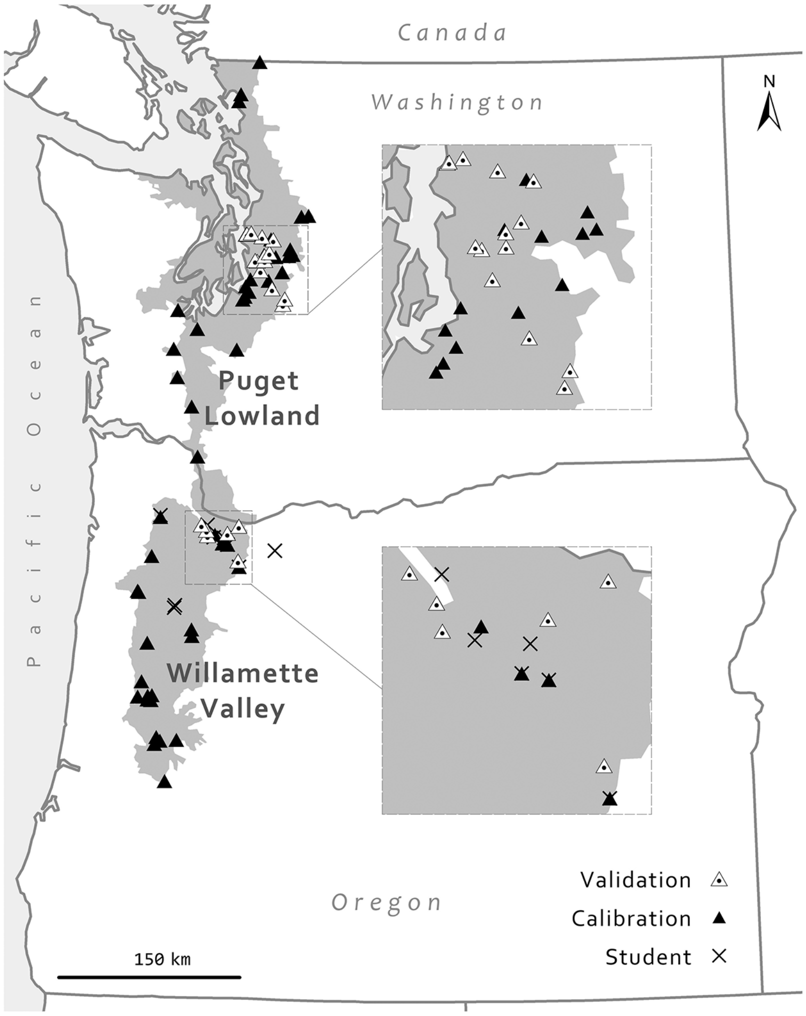

Figure 1. Map of benthic macroinvertebrate sampling locations in the Puget Lowland and Willamette Valley ecoregions of the Pacific Northwest, USA, for the development of a community science index of benthic integrity. Sampling was done for 3 separate datasets: validation, calibration, and student validation. Insets show enlarged portions of the landscape.

CS-IBI development and calibration

A continuously scaled macroinvertebrate IBI has not been developed for the Willamette Valley, so we used the Puget Lowland benthic IBI (B-IBI) as a framework for developing the CS-IBI (King County 2014). The B-IBI has been broadly applied in the Puget Lowland and is a strong indicator of watershed disturbance (Morley and Karr 2002, King County 2014). There are several advantages of using a professional, continuously scaled B-IBI as a framework for developing the CS-IBI, including a direct correspondence to a widely used professional IBI, the simplicity of metric development, and the ability to easily update metric values as more data are collected by community scientists. The B-IBI includes 10 macroinvertebrate metrics continuously scaled from 0 to 10 based on the observed 10^{th} and 90^{th} percentiles for each metric. The sum of the B-IBI metric scores, which range from 0 to 100, are used to categorize streams into 5 condition categories. In contrast with discrete metrics, continuously scaled metrics have less variability, a stronger association with environmental stressors, and a higher signal to noise ratio (Hughes et al. 1998, Blocksom 2003).

Calibration dataset and field methods To develop and calibrate the scoring criteria for the 10 metrics that make up the CS-IBI (Table 2), we generated a calibration dataset. Professional scientists collected the CS-IBI calibration samples from 17 August 2020 to 21 October 2020. They used a nonlethal method developed for community science (Edwards 2016) to collect, sort, subsample, identify, and count macroinvertebrates from 25 stream sites in the Puget Lowland and 23 sites in the Willamette Valley (48 total sites). One of the calibration sites (Clear Creek upper) was ~20 km east of the Willamette Valley ecoregion and at a slightly higher elevation than the other streams; however, we retained

Table 2. Description of metrics used in the community science index of biotic integrity (CS-IBI) for the Puget Sound Lowland and Willamette Valley ecoregions of the Pacific Northwest, USA, and the values for the 10th and 90th percentiles. If an observed value (OV) resulted in a score <0 or >10, a value of 0 or 10 was applied to that score. Scores for each metric are summed and range from 0 to 100, where a higher score indicates lower impairment. Expected relationship with watershed disturbance is indicated by + or –.

Metric	Description	10th	90th	CS-IBI formula
Taxa richness (–)	Macroinvertebrate families (no.)	8	15	10 × (OV – 8) / (15 – 8)
Mayfly richness (–)	Mayfly families (no.)	1	4	10 × (OV – 1) / (4 – 1)
Stonefly richness (–)	Stonefly families (no.)	0	3	10 × (OV – 0) / (3 – 0))
Caddisfly richness (–)	Caddisfly families (no.)	0	3	10 × (OV – 0) / (3 – 0)
Clinger richness (–)	Families that are clingers (no.)	3	9	10 × (OV – 3) / (9 – 3)
Long-lived richness (–)	Families that live longer than 1 y (no.)	0	2	10 × (OV – 0) / (2 – 0)
Intolerant richness (–)	Families that are pollution intolerant (no.)	0	1	10 × (OV – 0) / (1 – 0)
%dominant (+)	% of the top 3 most abundant taxa	61	89	10 × (OV – 61) / (89 – 61)
%predator (–)	% of organisms that are predators	0	12	10 × (OV – 0) / (12 – 0)
%tolerant (+)	% of organisms that are tolerant to pollution	0	25	10 × (OV – 0) / (25 – 0)

this site in the analysis because its watershed contained the lowest %Ag+Dev in our dataset.

To sample for this calibration dataset, the professional scientists used a 500-μm D-frame kick net (D-net) to collect a macroinvertebrate sample from the left, center, and right side of a riffle with cobble-sized substrate. They disturbed 0.09 m^2 of substrate in front of the D-net for ~1 min and gently rubbed the larger rocks with their hands. To avoid harming macroinvertebrates, they did not scrub the rocks with a brush or kick the substrate with their feet. The 3 riffle samples were composited for a total sample area of 0.27 m^2, sorted from debris, and poured into a divided plastic tray (38 × 24 × 6 cm, part #05905; Akro-Mils, Akron, Ohio) filled with stream water. Using a random number sheet, they selected ⅓ of the cells in the divided tray to randomly subsample 0.09 m^2 of benthic substrate. They used no-crush forceps, pipettes, and turkey basters to transfer macroinvertebrates into an ice-cube tray where they sorted, identified, and enumerated the subsample. Macroinvertebrates were identified to family level using a field guide specifically designed for in-the-field identification of live organisms (Edwards 2014). This process was repeated until at least 100 macroinvertebrates were counted and identified. The target count was based on the findings of Edwards (2016). To avoid bias, all organisms from the last randomly selected cell were counted (Edwards 2016). Because there was a range in the final counts for the calibration data, we rarefied the data to the minimum count (106) in the dataset. However, this step did not substantially change the metric formulas nor improve our statistical models, so we used the unrarefied data for analysis. All specimens were returned to the stream after identification and enumeration.

CS-IBI calibration We used the CS-IBI calibration dataset to calibrate the scoring criteria for each of the metrics in the B-IBI. Because of the coarser taxonomy of the community science data, we used 3 impairment categories instead of the 5 used in the B-IBI. The CS-IBI impairment categories are impaired (0–33), moderately impaired (34–66), and unimpaired (67–100). We assigned family-level traits for the CS-IBI data (Table S1) based on Fore et al. (2012) and determined the 10th and 90th percentiles for each metric (Table 2). If an observed value resulted in a metric score that was <0 or >10, we assigned a value of 0 or 10, respectively. We then calculated the CS-IBI score for each sample in the CS-IBI calibration dataset.

CS-IBI validation

To assess the accuracy of the CS-IBI scores, we used a validation dataset to independently validate the CS-IBI scores against scores from the Puget Lowland B-IBI. We also compared the individual metrics that make up the CS-IBI and B-IBI scores as well as their impairment categories.

Validation dataset The CS-IBI validation data were collected at 20 streams sites, with 14 in the Puget Lowland and 6 in the Willamette Valley (Fig. 1), by professional scientists from 6 August 2019 to 19 September 2019. At each 100-m reach, 2 riffles within 25 m of each other were concurrently sampled using a professional sampling method at 1 riffle and the community science nonlethal sampling method at the other riffle. Simultaneous sampling allowed for the direct comparison of the paired CS-IBI scores with professional IBI scores at each stream while minimizing spatial and temporal variance. The CS-IBI samples were collected as described above for the calibration dataset. The B-IBI samples were collected using a standard professional sampling technique (King County 2020). They used 500-μm D-nets (for Willamette Valley samples) or 500-μm Surber samplers (for Puget Lowland samples) to collect macroinvertebrates from 0.74 m^2 of substrate area in riffles by scrubbing the surface of rocks with brushes, followed by vigorously disturbing the substrate with a metal rod or by kicking the substrate by foot for 30 s. Samples were separated in the field and preserved in ethyl alcohol. Preserved samples were returned to the lab, and macroinvertebrates were subsampled from randomly selected grids in a Caton (1991) tray until a minimum count of 500 organisms or the entire sample was enumerated and identified. To minimize subsampling bias, all organisms from the last randomly selected grid were processed, even if the final count exceeded 500. The subsampled organisms were identified under magnification to the lowest practical level (usually genus or species) by an expert from a professional taxonomic company (Rhithron Associates, Missoula, Montana). We calculated the CS-IBI score for each sample in the CS-IBI validation dataset.

Validation To validate the CS-IBI, we compared 3 elements of the CS-IBI and B-IBI: individual metrics, impairment condition categories, and scores. There is no family-level equivalent of the B-IBI, thus, for these analyses, we compared the taxonomically coarser CS-IBI to the taxonomically finer B-IBI. First, we compared the individual metrics that make up the CS-IBI and B-IBI scores. For each of the 10 metrics that make up both IBIs (Table 2), we subtracted the CS-IBI score from the B-IBI score for each of the paired validation samples and summarized the difference in boxplots. Second, we evaluated the correspondence between the impairment condition categories of the paired CS-IBI and B-IBI scores. We aligned the B-IBI into 3 condition categories that matched the CS-IBI and compared the matching rate for each category, then used a χ^2 test in R (version 4.1.2; R Project for Statistical Computing, Vienna, Austria) to evaluate the similarity in impairment classifications between the indices. Third, we used Spearman's rank-based correlation test to test the relationship between CS-IBI and B-IBI scores.

CS-IBI reliability

Student dataset To test the reliability of the CS-IBI, we used a dataset of macroinvertebrates sampled by students under conditions encountered in the community science setting (Table 1). From 2005 to 2021, middle school, high school, and university students collected macroinvertebrates mainly in the autumn (September–November) or spring (May–June) during short field trips to 10 streams in the Willamette Valley ecoregion (90 total sampling events). These 10 streams were selected because we had a collection permit for them, and students could easily and safely access the stream sites. The selected reaches were generally representative of the overall stream conditions. Working together, groups of 3 to 5 students (mode = 4) spread out along a 100-m stream reach and used the nonlethal sampling method described above for the calibration dataset (Edwards 2016) to collect, sort, subsample, identify, and count macroinvertebrates. Students were briefly trained in the nonlethal method before the field trip and then again during the field trip, where they were able to watch an in-field demonstration. Students were shown an example of a riffle (vs a pool or glide) and then directed to select a riffle from which they collected macroinvertebrates. The maximum count of macroinvertebrates collected at a stream during a sampling event was determined by the number of students on the field trip and, thus, could not be controlled. Many of the field trips consisted of multiple classes collecting macroinvertebrates at the same stream resulting in a large area of substrate sampled for some streams (Table 3). The final macroinvertebrate counts and identifications of each group were reviewed by someone who was familiar with the macroinvertebrate identification. We (the lead and 2nd authors of this paper) verified ~50% of the samples, and the other ~50% of the samples were verified by college students or high school teachers who were not formally trained in macroinvertebrate identification but had previously attended a field trip and were familiar with common macroinvertebrate families. We calculated the CS-IBI score for each sample in the student dataset.

Variability We used a subset of the student dataset to estimate the interannual and seasonal variability in CS-IBI scores from Balch Creek and Rock Creek, which were repeatedly sampled in the spring and autumn by students from 2005 to 2020 (Table 3) and had relatively high CS-IBI scores. Balch Creek was sampled 31 times (spring = 15, autumn = 16), and Rock Creek was sampled 17 times (spring = 8, autumn = 9). We estimated the seasonal interannual variance by determining the absolute difference in the year-to-year CS-IBI scores for the spring and autumn samples (e.g., |autumn 2019–autumn 2020|). We used the 95% CI of the absolute differences in CS-IBI scores for each season to estimate the variability of CS-IBI scores generated by community scientists.

Comparing macroinvertebrate assemblages

We compared macroinvertebrate assemblage data generated by the nonlethal community science method with assemblage data generated by standard macroinvertebrate sampling methods. We did not have information about student macroinvertebrate misidentification rates, though this issue has been investigated in previous studies (Fore et al. 2001, Engel and Voshell 2002, Nerbonne and Vondracek 2003, Edwards 2016). Instead, we compared the assemblage data generated by both methods by aligning the B-IBI and CS-IBI taxonomy in the validation dataset and comparing the resulting assemblages based on macroinvertebrate relative abundance and ordinations. To align the taxonomy, we reclassified the B-IBI data to the family, order, or class level that matched the CS-IBI taxonomy. Using boxplots, we compared the relative abundance of each taxon collected in the validation data and selected taxa with >20% difference in relative abundance for further evaluation.

Table 3. Summary data for the student dataset collected by middle school, high school, and university students from 2005 to 2021 in 10 streams in the Willamette Valley ecoregion, Pacific Northwest, USA. A sampling event represents a field trip to a stream in a given season (autumn or spring) and year, except for Balch Creek, which was sampled 3 times in the winter (December). In many cases, >1 sampling event occurred per season, and these samples were aggregated. The mean macroinvertebrate counts represent the total number of macroinvertebrates subsampled and counted during a sampling event averaged across all events for that stream. Mean stream area sampled and mean community science index of biotic integrity (CS-IBI) scores are provided for each stream. Summary statistics, including 1 SD are provided for streams with >1 sampling event. NA indicates streams with only 1 sampling event. Max = maximum, min = minimum.

Stream	Student level	Sampling events	Mean macroinvertebrate counts (min–max)	Mean area (m^2) sampled (min–max)	Mean CS-IBI (SD)
Balch	University	34	384 (107–1164)	1.1 (0.5–3.3)	62.1 (13.9)
Carli	High school	10	293 (97–903)	2.4 (0.6–5.0)	14.5 (1.9)
Clark	Middle school	1	109 (NA)	0.7 (NA)	13.5 (NA)
Clear (Lower)	University	5	106 (77–186)	0.7 (0.6–0.8)	62.9 (16.1)
Clear (Upper)	University	2	170 (83–256)	1.1 (0.6–0.8)	77.5 (5.8)
Gales	High school	3	140 (101–179)	0.8 (0.2–1.5)	77.9 (2.6)
Johnson	High school	5	174 (57–307)	1.7 (0.6–2.5)	20.7 (7.3)
Mt Scott	High school	12	507 (93–1320)	2.3 (0.6–4.1)	34.6 (13.2)
Pringle	Middle school	1	108 (NA)	0.8 (NA)	18.3 (NA)
Rock	High school	17	875 (104–3688)	2.9 (0.7–7.2)	71.3 (10.5)

We used ordination, a data visualization technique in which samples are plotted along 2 or more axes (Legendre and Legendre 1998), to summarize the structure of the macroinvertebrate assemblage data and evaluate differences in the assemblages between the CS-IBI and B-IBI validation data. In assemblage-based ordinations, points that are closer together are more similar in assemblage than points that are far apart. We used nonmetric multidimensional scaling (NMDS) based on Bray–Curtis distance (McCune and Grace 2002) in the *vegan* package (version 2.6-7; Oksanen et al. 2020) to ordinate assemblage data (20 random starts) and used Procrustes analysis to compare the similarity of the CS-IBI and B-IBI data. Procrustes rotates the ordinations to best match the paired samples and uses the M^2 statistic and the Protest permutation-based test (permutations = 999) to evaluate the similarities of the ordinations (Jackson 1995, Peres-Neto and Jackson 2001). The Procrustes M^2 statistic ranges from 0 to 1, with larger values indicating a stronger correspondence between the paired samples in ordination space (Jackson 1995). We expected that the B-IBI taxonomically aligned assemblage data generated by professional scientists would be similar to the CS-IBI assemblage data generated by community scientists using the nonlethal field method. There were 4 creeks that had a relatively large difference between CS-IBI and B-IBI locations in ordination space (Johnson, Thornton, Evans, and Kelley creeks; see Results). To further evaluate the assemblage differences in these 4 streams, we compared the relative abundance of the 10 most abundant taxa in each stream using stacked bar plots.

Relationships with land-use stressors

To evaluate the response of the CS-IBI to watershed stressors, we fit simple linear models and r^2 values to characterize the relationships between %Ag+Dev in the watershed and 1) CS-IBI calibration scores, 2) CS-IBI and B-IBI validation scores, and 3) student CS-IBI validation scores. After we fit the linear regression models, we performed a model diagnostic check on the residuals. We used a Shapiro–Wilk test to test the normality assumption and the F-ratio to test the equal variance assumption, which all datasets passed. For the calibration and validation datasets, we fit models for the entire geographic range and for each ecoregion separately. We expected B-IBI and CS-IBI scores to decrease with %Ag+Dev in the watershed in all cases. Because many of the streams in the student validation dataset were repeatedly sampled (Table 3), we first used the mean CS-IBI sores and then repeated the analysis with 5 randomly selected sets of samples from each stream that had >1 sampling date. For streams sampled only once, we used the data from the single available sample for that stream. To further evaluate CS-IBI scores generated by participants with minimal experience identifying macroinvertebrates, we also fit a linear model with only the data from 6 streams in which the macroinvertebrate identifications and counts were verified by relatively inexperienced teachers or college students.

RESULTS

CS-IBI development and calibration

We collected 48 calibration samples across the 2 ecoregions (Fig. 1). The formula for each CS-IBI metric can be found in Table 2. Mean %Ag+Dev in the watershed for the calibration data was 28% (range = 0–95%). Mean abundance of subsampled organisms was 161 (range = 106–296), and mean area sampled was 0.11 m^2 (range = 0.09–0.22 m^2). The mean CS-IBI score was 51 (range = 12–84), where a lower score indicates greater impairment. Fifteen CS-IBI scores were in the impaired category (<33), 17 CS-IBI scores were in the moderately impaired category (33–66), and 16 scores were in the unimpaired category (>66).

CS-IBI validation

We collected 20 paired validation samples across both ecoregions (Fig. 1). Mean %Ag+Dev in the watershed for the validation samples was 54% (range = 1–87%). Mean abundances of subsampled organisms for the B-IBI data and CS-IBI data were 483 and 154, respectively. The mean B-IBI score was 35 (range = 7–95), and the mean CS-IBI score was 39 (range 10–82). Of the 10 metrics that make up the CS-IBI, 1 metric was consistently overestimated (%intolerant), 2 were slightly overestimated (caddisfly richness and %dominant), and 1 was underestimated (%clinger) by the nonlethal community science method (Fig. 2A). The mean difference in the absolute scores for all metrics was 2.8 (range = 2.1–4.8). The metrics with the largest mean difference between the CS-IBI and B-IBI absolute scores were %tolerant (mean difference = 4.8) and %dominance (mean difference = 3.9). The mean difference in the absolute scores of the other 8 metrics was <3.0.

Of the 20 paired samples, the CS-IBI categorized 11, 6, and 3 of the streams as impaired, moderately impaired, and unimpaired, respectively. The B-IBI categorized 13, 3, and 4 of the streams as impaired, moderately impaired, and unimpaired, respectively. Results of the χ^2 analysis showed that the CS-IBI and B-IBI condition categories were similar (χ^2 = 6.0, df = 4, p = 0.20), and 13 of the 20 condition categories matched. Four of the mismatched classifications occurred in the impaired to moderately impaired categories, with differences between CS-IBI scores and B-IBI scores of 7, 26, 33, and 37. Three of the mismatched classifications occurred in the moderately impaired to unimpaired categories, with differences in scores of 18, 15, and 40. Of the 7 mismatches, none diverged >1 category in difference.

CS-IBI scores were correlated with B-IBI scores (ρ = 0.62, p < 0.01). There were 4 streams with mismatched

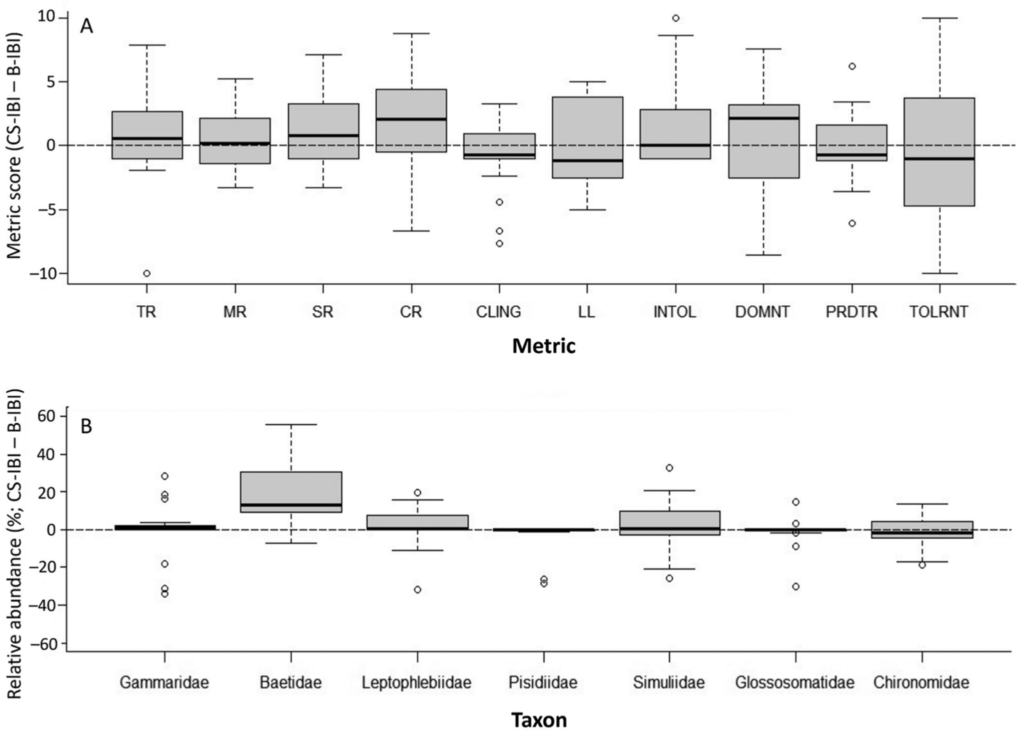

Figure 2. Boxplots comparing the differences between individual metrics (A) and macroinvertebrate relative abundance (B) for the community science index of biotic integrity (CS-IBI) and benthic IBI (B-IBI) paired samples from the validation dataset. Boxes encompass the lower quartile, median, and upper quartile of the data. Whiskers extend to 1.5× the interquartile range, and the open circles are 1.5× the interquartile range. Values above and below the dotted line indicate an overestimate or underestimate by the CS-IBI, respectively. Individual metrics in panel A are taxa richness (TR), mayfly richness (MR), stonefly richness (SR), caddisfly richness (CR), clinger richness (CLING), long-lived richness (LL), intolerant richness (INTOL), %dominant (DOMNT), %predator (PRDTR), and %tolerant (TOLRNT). See Table 2 for details about how each metric is calculated. Panel B shows only taxa with maximum differences in relative abundance of >20% between the CS-IBI and B-IBI datasets.

condition categories and absolute differences of >20 points between the CS-IBI and B-IBI total scores: Evans, Panther, Vasa, and Taylor creeks. In these streams, intolerant richness and %predator were the metrics that showed the largest and most frequent source of variation in scores, with the CS-IBI overestimating B-IBI scores in every case.

Student CS-IBI

The student dataset (Table 3) was generated by >6000 students collecting macroinvertebrates from 10 streams during 90 sampling events from 2005 to 2021 (autumn = 52, spring = 33, other season = 5). The mean area sampled in each stream ranged from 0.7 to 2.9 m^2 at each stream. The mean CS-IBI scores for each stream ranged from 13.5 to 77.9. At Balch Creek, mean interannual seasonal variability of the student CS-IBI scores was 12.8 (95% CI = 2.3–23.3) in the autumn and 12.7 (95% CI = 4.2–21.3) in the spring. At Clear Creek, mean interannual variability was 7.8 (95% CI = 0.9–14.7) in the autumn and 16.7 (95% CI = 6.0–27.5) in the spring. The mean variability of CS-IBI scores for both streams was 10.3 in the autumn and 14.8 in the spring. The mean 95% CI for both streams was 9.0 in the autumn and 16.2 in the spring.

Comparing macroinvertebrate assemblages

After taxonomic alignment, total richness of the CS-IBI and B-IBI validation data was 49. Six taxa had maximum differences in relative abundance >20% between the CS-IBI and B-IBI datasets (Fig. 2B). Of these 6 taxa, Baetidae was consistently overestimated, and Pisidiidae was consistently underestimated by the CS-IBI. The other 4 taxa showed similar variation but were not biased towards either method. The nonmetric multidimensional scaling ordinations (k = 2; Fig. S1) generated by CS-IBI and B-IBI data were similar (Procrustes $M^2 = 0.57, p = 0.002$; Fig. 3). The relatively large differences in assemblage between the taxonomically aligned CS-IBI and B-IBI data in 4 creeks (Johnson, Thornton, Evans, and Kelley) were primarily driven by small taxa such as Elmidae and Pisidiidae, which were frequently missed during field sorting, as well as taxa that adhere to the substrate, such as Glossosomatidae and Hydroptilidae. Baetidae and Simuliidae were consistently overestimated in the CS-IBI data (Fig. 4).

Relationships with land use

The calibration, validation, and student datasets were related to watershed land use (Table S2). CS-IBI calibration scores had a negative relationship with %Ag+Dev in the watershed ($y = -0.50x + 64.2$, $r^2 = 0.56$, $p < 0.001$; Fig. 5A). Within ecoregions, the relationship with %Ag+Dev was stronger for the Puget Lowland ($r^2 = 0.71$, $p < 0.001$) than for the Willamette Valley ($r^2 = 0.45, p < 0.001$). B-IBI validation scores decreased with %Ag+Dev land use in the watershed ($y = -0.77x + 76.2, r^2 = 0.74, p < 0.001$; Fig. 5B), as did

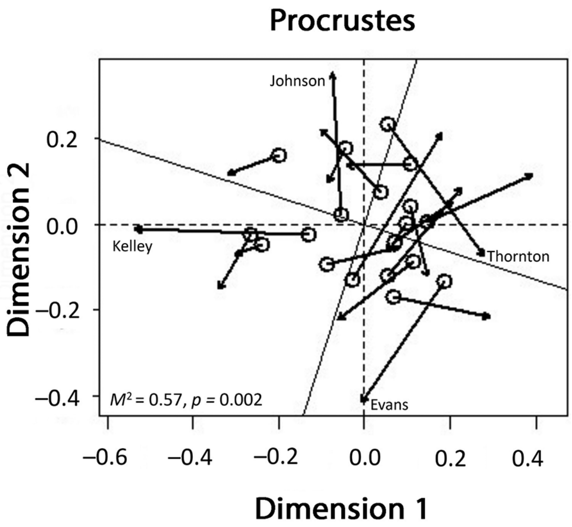

Figure 3. Procrustes analysis of separate nonmetric multidimensional scaling (NMDS) ordinations of Bray–Curtis dissimilarity matrices of assemblage relative abundance data collected by a nonlethal sampling method for a community science index of biotic integrity (CS-IBI) or a professional sampling method for a benthic IBI (B-IBI). The ordination shows the locations of the paired CS-IBI samples and the B-IBI samples from the taxonomically aligned validation data in ordination space (k = 2). Circles are the B-IBI samples, and the arrows point to the corresponding CS-IBI samples. The solid lines indicate the rotation necessary to align the CS-IBI NMDS ordination with the B-IBI NMDS. Four creeks that had a relatively large difference between CS-IBI and B-IBI locations in ordination space (Johnson, Thornton, Evans, and Kelley creeks) are noted. M^2 is a measure of similarity between ordinations.

CS-IBI validation scores within each watershed ($y = -0.52x + 66.8$, $r^2 = 0.50$, $p < 0.001$; Fig. 5C). Likewise, student-collected mean CS-IBI scores decreased with %Ag+Dev in the watershed ($y = -0.70x + 81.3$, $r^2 = 0.77$, $p < 0.001$; Fig. 5D). The mean r^2 value of the linear models for the 5 randomly selected student CS-IBI scores was 0.68 (r^2 range = 0.38–0.85). For the 6 streams in which the student counts and identifications were identified by inexperienced volunteers, the r^2 value of the linear model was 0.55 ($p = 0.09$).

DISCUSSION

The main objectives of this study were to develop and validate an IBI for community science that corresponded with a professional IBI and could be applied in the Willamette Valley and Puget Lowland ecoregions, wherein lie the majority of community science groups and watershed councils in the PNW. Validating and assessing the variability of the CS-IBI has provided an IBI that is both valid and reliable for the Willamette Valley and Puget Lowland ecoregions and that reflects the relationship between stream integrity and the amount of development and agriculture in the watershed.

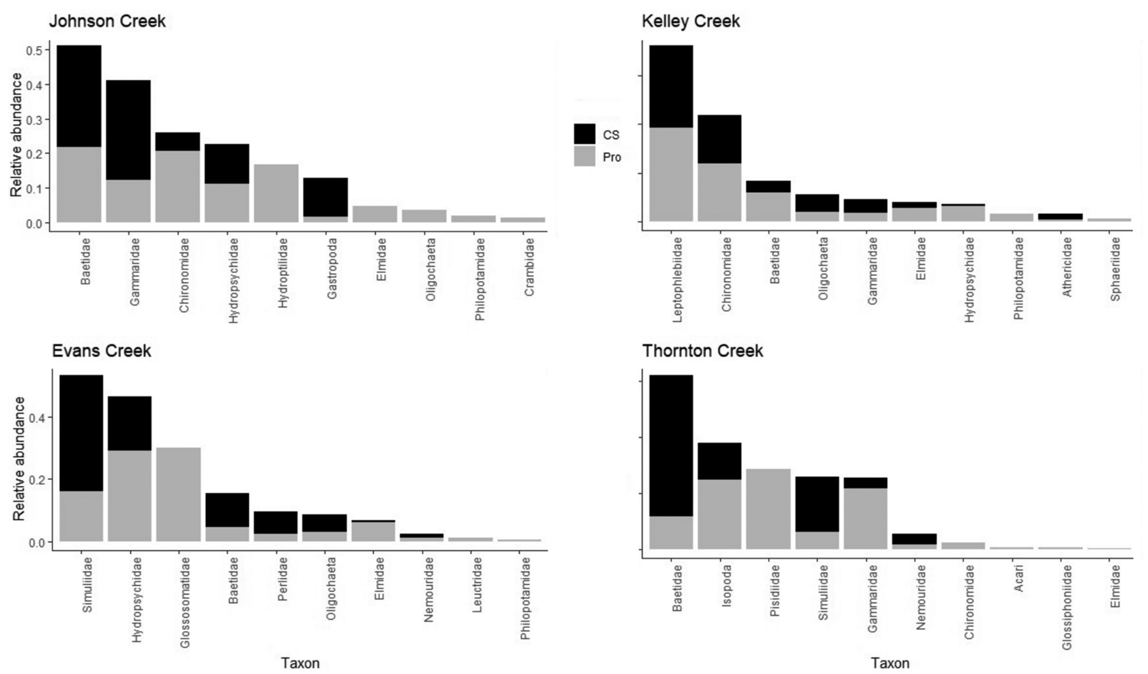

Figure 4. Relative abundance data for the top 10 most common taxa at 4 creeks that showed relatively large differences in assemblage. Data for the paired samples were generated with a nonlethal community science (CS) sampling method or a professional (Pro) sampling method.

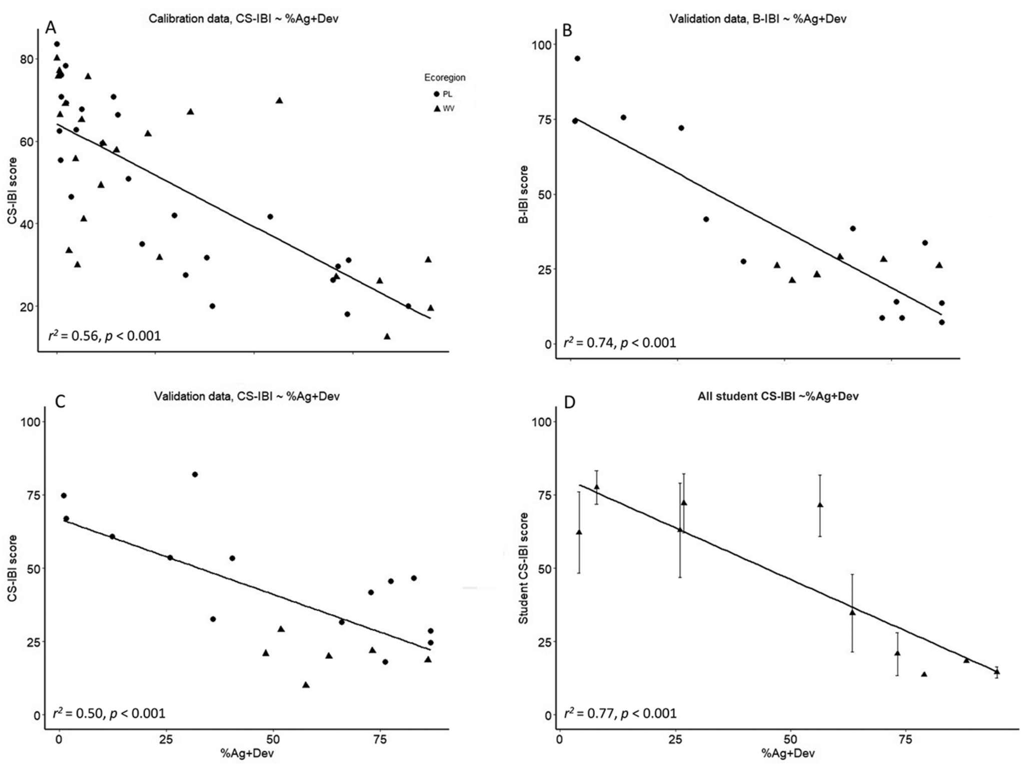

Figure 5. Scatterplots of the calibration and validation data and results of linear models with each dataset as a function of the percentage of agriculture + developed land use (%Ag+Dev) in the watershed: the community science index of biotic integrity (CS-IBI) calibration data (A), the professional benthic IBI (B-IBI) validation data (B), the CS-IBI validation data by ecoregion (C), and the student CS-IBI scores (D), with the error bars representing 1 SD. PL = Puget Sound Lowland, WV = Willamette Valley.

The CS-IBI is a useful tool for stream managers to evaluate stream condition and engage the public to support comprehensive stream management plans.

CS-IBI validation

We report here the 1st study to calibrate and estimate the variability of a trait-based IBI for community science that corresponds to, and is validated with, a professional IBI. The correspondence between the CS-IBI and B-IBI 3-class stream impairment condition categories had a match rate of 65%. This finding aligns with those of Pinto et al. (2020), who found a match rate between a community science metric and a professional metric of 49 to 58% for a 5-class condition category and 85 to 91% for a 2-class condition category. In addition, the correlation between CS-IBI and B-IBI scores ($\rho = 0.62$) is similar to previous findings. For example, Moffett et al. (2015) found that scores generated by community scientists were correlated with professional scores in New Zealand streams ($\rho = 0.54$, $p < 0.001$), and O'Leary et al. (2004) found that professional genus- and family-level scores correlated with several different metrics (ρ range $= 0.32$–0.78) in streams of the northeast United States. Fore et al. (2001) found much stronger relationships between community science data and professional data ($\rho = 0.98$, $p < 0.01$); however, their study controlled for sorting method and within-reach sample location.

Differences between the CS-IBI and B-IBI scores were primarily driven by high variability in the total richness, %tolerant, and %dominant metrics, but differences were also associated with bias of the nonlethal sampling method. This source of error in the %tolerant metric was due to the relatively few macroinvertebrate families categorized as intolerant ($n = 2$) or tolerant ($n = 4$). The variability in the total richness and %dominant metrics could have been due to the more accurate sorting of samples in the lab under magnification vs field-based identification, as well as the finer taxonomy used in the professional samples. The underestimation of clinger richness was likely due to differences in how aggressively professionals scrub the rocks and disturb the substrate compared with community scientists, who may be trying not to harm organisms.

Differences in assemblage between the CS-IBI and B-IBI samples can also be generally explained by the limitations of the nonlethal method, including the challenge of identifying small organisms in the field with only hand lenses and under ambient light conditions. These conditions tend to favor the identification of relatively larger macroinvertebrates that swim or actively move in the collection trays and, thus, are easier to see. This source of bias has been documented previously in studies of other live-sort methods (Nerbonne et al. 2008, Edwards 2016). In the 4 streams that showed large differences in assemblage between the CS-IBI and B-IBI data (Fig. 4), there was high variation in the relative abundance of Baetidae and Gammaridae, likely because these families rapidly swim around the sampling tray and are more likely to be seen during the field sort. Similarly, Simuliidae were overrepresented in community science samples, likely because these larvae attach to the sides of the ice-cube tray, making them easier to see with the naked eye. There was also high variation in small nonmotile taxa, such as Pisidiidae, that are easily obscured in debris. Despite variation and bias in the macroinvertebrate relative abundance data generated by community scientists, the resulting assemblages were statistically similar to the professional data.

The seasonal interannual mean variability in student CS-IBI scores (range $= 7.8$–16.7) was similar to what has been observed in B-IBI scores (8.7) generated by professional scientists (CAL, unpublished data); however, the 95% CI of variation in the student-generated CS-IBI scores (range $= 14.7$–27.5) was much higher. This difference was likely due to higher variability and coarser taxonomic resolution of the nonlethal method, but it also may be due to the substantially longer time-period and more numerous samples of the Balch Creek ($n = 31$) and Rock Creek ($n = 17$) data as compared with the unpublished CAL data ($n = 2$–6 per stream). In addition, mean variability in the student CS-IBI scores had a seasonal pattern, with lower variability in autumn than spring. This seasonality was likely related to variability in streamflow conditions, which tend to be more stable in the late summer and early autumn during baseflow conditions in the PNW. Furthermore, because of the timing of flows in the PNW where streamflow can be too high for sampling from November to June, most community science would take place in midsummer to late autumn, which is within the time period that the calibration samples were collected. In the student data, nearly ⅔ of all samples were collected in the autumn. The relatively high variability observed in the Rock Creek data was most likely due to major stream habitat restoration that occurred in spring 2014 and resulted in post-restoration macroinvertebrate communities that were highly variable (Bedell 2015, Edwards et al. 2018).

Relationships of IBI scores with watershed land-use stressors

The CS-IBI performed well in identifying the relative biotic integrity of streams, as evidenced by its score's relationship with the amount of development and agriculture in the watershed. Scores from the CS-IBI calibration data, CS-IBI and B-IBI validation data, and student-generated CS-IBI were negatively related to the %Ag+Dev in the watershed, showing a clear response with increasing anthropogenic stressors. Our results were similar to what has been observed in other studies of Ephemeropteran, Plecopteran, and Trichopteran richness and land use in the Willamette Valley ($R^2 = 0.61$–0.71; Van Sickle et al. 2004, Waite et al. 2010) and the B-IBI and land use in the Puget Lowland ($\rho = 0.69$; King County 2014). Compared with the B-IBI, the calibration and validation CS-IBI scores had a weaker relationship with watershed stressors, which is likely due to differences in taxonomic resolution (Chessman et al. 2007), increased variability of the nonlethal method (Nerbonne et al. 2008), and the fewer number of macroinvertebrates that were subsampled and counted (Doberstein et al. 2000). Despite the weaker relationship between validation CS-IBI validation scores and %Ag+Dev in the Willamette Valley, student CS-IBI scores were strongly associated with watershed land use in the Willamette Valley. The relatively weak evidence for a relationship between CS-IBI scores and %Ag+Dev for data that were verified by inexperienced teachers or students was likely due to low sample size ($n = 6$). However, the r^2 value was similar to that of the CS-IBI validation data, implying that even in this smaller subset of data with less certain taxa identification, the CS-IBI generally captured the same pattern in biotic integrity along the gradient of watershed development.

Incorporating the CS-IBI into comprehensive stream management

Although the precision of the CS-IBI is not high enough for resource managers to use for regulatory purposes, the similarity between the B-IBI and CS-IBI in diagnosing watershed condition suggests the CS-IBI would be useful to resource managers in several ways. First, the CS-IBI could be used as a screening tool, where results from the CS-IBI could be used to set priorities for their limited professional-level monitoring funds. For example, correspondence between CS-IBI and B-IBI impairment categories was always within 1 category, suggesting that the CS-IBI reliably detects a change of >1 category. The mean interannual 95% CI of CS-IBI scores in the student-generated data was within 1 impairment category (i.e., 33 points), so if CS-IBI scores for a stream were more than the mean 95% CI (9.0–16.2), stream managers could initiate more in-depth monitoring or professional assessment. The reliability of the CS-IBI would be particularly useful for management plans that use periodic sampling designs with multiyear gaps between professional assessments because an observed change >1 in the CS-IBI impairment category could signal the immediate need for professional bioassessment. Next, the seasonal differences in variability that we observed could help stream managers determine which season to conduct community science bioassessment. For example, our findings suggest that community science groups may want to sample in the summer or autumn during low-flow conditions, when the CS-IBI variability is relatively low and stream conditions are safer for sampling by nonprofessionals. Finally, a substantial advantage of the nonlethal method and the CS-IBI is their ability to deliver immediate results in contrast with professional data, which can be costly to generate and often require several months of lab work before results can be obtained. These rapid results facilitate timely response to a pollution event and allow for the rapid bioassessment of streams, which will help managers make prompt and informed decisions when prioritizing environmental management efforts.

There are several important considerations for incorporating community science and the CS-IBI into comprehensive stream and watershed management plans. The CS-IBI should only be applied to macroinvertebrate samples collected from riffles in flowing streams with rocky substrate. Slow water environments with sand or silt substrate cannot be reliably evaluated with the CS-IBI. Furthermore, the CS-IBI should not be used to score samples with <100 organisms, samples sorted under high magnification, or samples that were nonrandomly subsampled. Finally, to ensure the macroinvertebrate data are accurate, it is important to have a trained participant who can verify community scientists' macroinvertebrate counts and identifications before specimens are returned to the stream.

A place for community science in stream management

Rieman et al. (2015) argue that stream managers must engender broad public support, otherwise management actions are more difficult to implement. Among many recommendations, Rieman et al. (2015) suggest that public education and community science monitoring should be central elements of comprehensive management plans. In contrast, skepticism by experts about the quality of data collected by volunteers limits its use in environmental management (Penrose and Call 1995, Bonney et al. 2014). We believe that the development of reliable bioassessment tools for community science supports broader efforts to engage the public in environmental management because the perception by participants that their data are being used to solve an environmental problem is an important factor in motivating participation in community science (Nerbonne et al. 2008, Phillips et al. 2019, Larson et al. 2020). The CS-IBI is based on a professional IBI and calibrated for use in community science; thus, it raises the value of such data to environmental managers while increasing the likelihood that the public will continue to participate in community science programs. Furthermore, when students participate in authentic research experiences, they are more likely to select science, technology, engineering, and math majors and persist in their educational pursuits (NASEM 2018). These reciprocal benefits directly support efforts to increase the social capital of environmental management and broaden the societal impacts related to stream restoration (Dickinson et al. 2012).

One of the primary objectives of community science data is to fill the gap that exists between management agencies' monitoring programs and the funding resources that would be required to gather the data at all local spatial scales across a state. To that end, states often provide a substantial level of funding to community science groups to do local monitoring, which is in turn used by the states to meet management goals. In our experience, states that support and use community science also require training, oversight, and controls to ensure community science data meet quality assurance objectives. These safeguards may come in the form of direct state staff oversight, but, more frequently, community science groups partner with experienced or trained staff to lead inexperienced community members in monitoring activities. The methods used in this study require similar oversight to all other community science data used by Oregon and Washington state resource agencies: a person, with the appropriate education or training, leads a group of less-qualified individuals in data collection, followed by review and, if necessary, corrections to the data.

We acknowledge that not all community science programs verify identifications made by community scientists (i.e., Pinto et al. 2020), and we recognize the challenge presented by the need for verification of macroinvertebrate identifications. However, the inclusion of a returning student, teacher, or other volunteer to verify macroinvertebrate identifications can provide a culturally familiar role model and deepen the experience for participants. Students who interact with a peer or mentor from a familiar culture and similar demographic background show increased persistence and engagement with science (Tsui 2007). We feel that assuring data quality and deepening the social experience for community scientists justifies the effort to recruit experienced volunteers to verify identifications made by community scientists.

It is important to note that we are not advocating for the CS-IBI to be used for regulatory purposes. Instead, we envision explicitly linked programs between environmental agencies and community science groups to generate supplementary data as part of a comprehensive watershed management plan. As a tool, the CS-IBI enables further collaboration between environmental managers and the communities they serve, tapping into the potential value of community science to improve stream health. We believe that the CS-IBI has several advantages as a bioassessment tool, including correspondence with a professional index, ease of development and calibration, immediate results, and intuitive scoring that is easily understood by the public. We believe that these advantages increase the value of community science data and will help broaden efforts to monitor and restore stream ecosystems. We hope our approach serves as a model for how agencies can partner with community groups to develop and validate a community science IBI for their region.

ACKNOWLEDGEMENTS

Author contributions: PME, SLH, CAL, KHM, and JW conceived, designed, and implemented the study. PME, DB, CAL, SLH, EM, and KHM wrote the manuscript. PME, DB, SLH, KHM, and CP collected data. PME, DB, SLH, CAL, KHM, EM, CP, EW, and JW designed components of the study and contributed to data analysis. PME, DB, SLH, CAL, KHM, EM, CP, EW, and JW reviewed and edited the manuscript.

We would like to acknowledge and thank students at Clackamas High School, Milwaukee High School, Rex Putnam High School, Jane Goodall Environmental Middle School, and Portland State University, who collected data. We also thank teachers Christopher Downie, Leigh Grange, Doug Gunter, Phil Gwin, Phil Marchant, Jordon Nuss, Angie Shroufe, and Rod Shroufe for organizing field trips for their students. We would also like to thank Megan Colley, Lex Francis, Dakota Hufford, Shauna Lodi, and Hannah Smiley for their help on field trips and Emily Rahlmann for help sampling at the validation sites. We thank Dr Yangdong Pan's Water Quality lab group for reading and commenting on an early draft of the manuscript. We also thank Zee Searles Mazzacano for helping interpret community science data.

This study was funded by the Environmental Professional Program at Portland State University. Data collected by students were funded by Clackamas County Water Environment Services, Hillsboro Clean Water Services, and the Student Watershed Research Project at Portland State University.

LITERATURE CITED

Aukland Council. 2013. Wai Care. (Available from: https://waicare.org.nz/Home.aspx)

Ballard, H. L., C. G. Dixon, and E. M. Harris. 2017. Youth-focused citizen science: Examining the role of environmental science learning and agency for conservation. Biological Conservation 208:65–75.

Bedell, D. 2015. Rock Creek restoration macroinvertebrate monitoring report. Environmental Science and Management Professional Master's Project Reports. Portland State University, Portland, Oregon. (Available from: https://doi.org/10.15760/mem.45)

Blocksom, K. A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. Environmental Management 31:670–682.

Bonney, R. 2021. Expanding the impact of citizen science. BioScience 71:448–451.

Bonney, R., C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk. 2009. Citizen science: A developing tool for expanding science knowledge and scientific literacy. BioScience 59:977–984.

Bonney, R., T. B. Phillips, H. L. Ballard, and J. W. Enck. 2016. Can citizen science enhance public understanding of science? Public Understanding of Science 25:2–16.

Bonney, R., J. L. Shirk, T. B. Phillips, A. Wiggins, H. L. Ballard, A. J. Miller-Rushing, and J. K. Parrish. 2014. Next steps for citizen science. Science 343:1436–1437.

Brown, E. D., and B. K. Williams. 2019. The potential for citizen science to produce reliable and useful information in ecology. Conservation Biology 33:561–569.

Burdett, A. S., K. E. O'Reilly, R. J. Bixby, and S. S. Connealy. 2021. How to get your feet wet in public engagement: Perspectives from freshwater scientists. Freshwater Science 40:228–237.

Callaghan, C. T., J. J. Rowley, W. K. Cornwell, A. G. Poore, and R. E. Major. 2019. Improving big citizen science data: Moving beyond haphazard sampling. PLoS Biology 17:e3000357.

Caton, L. W. 1991. Improving subsampling methods for the EPA's "Rapid Bioassessment" benthic protocols. Bulletin of the North American Benthological Society 8:317–319.

Chessman, B., S. Williams, and C. Besley. 2007. Bioassessment of streams with macroinvertebrates: Effect of sampled habitat and taxonomic resolution. Journal of the North American Benthological Society 26:546–565.

Davies, S. P., and S. K. Jackson. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16:1251–1266.

Dewitz, J. 2019. National Land Cover Database (NLCD) 2016 products. United States Geological Survey data release. (Available from: https://doi.org/10.5066/P96HHBIE)

Dickinson, J. L., J. Shirk, D. Bonter, R. Bonney, R. L. Crain, J. Martin, T. Phillips, and K. Purcell. 2012. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment 10:291–297.

Doberstein, C. P., J. R. Karr, and L. L. Conquest. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshwater Biology 44:355–371.

Edwards, P. M. 2014. Stream insects field guide. Center for Science Education, Portland State University, Portland, Oregon. (Available from: https://www.pdx.edu/environmental-science/field-guide)

Edwards, P. M. 2016. The value of long-term stream invertebrate data collected by citizen scientists. PLoS One 11:e0153713.

Edwards, P. M., and G. Shaloum. 2020. Citizen science and ecological restoration. Society for Ecological Restoration News 34:16–18.

Edwards, P. M., G. Shaloum, and D. Bedell. 2018. A unique role for citizen science in ecological restoration: A case study in streams. Restoration Ecology 26:29–35.

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

Firehock, K. 1994. Save our streams: Volunteer trainer's handbook. Izaak Walton League of America, Gaithersburg, Maryland.

Firehock, K., and J. West. 1995. A brief history of volunteer biological water monitoring using macroinvertebrates. Journal of the North American Benthological Society 14:197–202.

Fore, L., R. Wisseman, J. O. Wilhelm, D. Lester, K. Adams, G. Hayslip, and P. Leinenbach. 2012. Using natural history attributes of stream invertebrates to measure stream health. King County Department of Natural Resources and Parks, Seattle, Washington. (Available from: http://pugetsoundstreambenthos.org/Projects/EPA_Grant_2010/TechDocs/Final/Invertebrate_Attributes_2013_v2.pdf)

Fore, L. S., K. Paulsen, and K. O'Laughlin. 2001. Assessing the performance of volunteers in monitoring streams. Freshwater Biology 46:109–123.

Fraisl, D., G. Hager, B. Bedessem, M. Gold, P. Y. Hsing, F. Danielsen, C. B. Hitchcock, J. M. Hulbert, J. Piera, H. Spiers, and M. Thiel. 2022. Citizen science in environmental and ecological sciences. Nature Reviews Methods Primers 2:64.

Hopfensperger, K. N., E. I. Larson, S. E. Washko, and E. K. Moody. 2021. Elevate your work through incorporation of public engagement. Freshwater Science 40:221–227.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, T. M. Kincaid, L. Reynolds, and D. P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences 55:1618–1631.

Jackson, D. A. 1995. PROTEST: A PROcrustean Randomization TEST of community environment concordance. Ecoscience 2:297–303.

Karr, J. R. 1998. Rivers as sentinels: Using the biology of rivers to guide landscape management. Pages 502–528 *in* R. J. Naiman and R. E. Bilby (editors). River ecology and management: Lessons from the Pacific coastal ecoregion. Springer, Berlin, Germany.

Karr, J. R., and E. W. Chu. 2000. Sustaining living rivers. Hydrobiologia 422:1–14.

King County. 2014. B-IBI restoration decision framework and site identification. Prepared by Jo Opdyke Wilhelm, Debra Bouchard, Chris Gregersen, Chris Knutson, and Kate Macneale. Water and Land Resources Division, Seattle, Washington. (Available from: https://pugetsoundstreambenthos.org/Projects/Restoration-Priorities-2014.aspx)

King County. 2020. King County freshwater benthic macroinvertebrate sampling and analysis plan. Prepared by Kate Macneale, Liora Llewellyn, and Beth Sosik. Water and Land Resources Division, Seattle, Washington. (Available from: https://pugetsoundstreambenthos.org/Projects/Restoration-Priorities-2014.aspx)

Larson, L. R., C. B. Cooper, S. Futch, D. Singh, N. J. Shipley, K. Dale, G. S. LeBaron, and J. Y. Takekawa. 2020. The diverse motivations of citizen scientists: Does conservation emphasis grow as volunteer participation progresses? Biological Conservation 242:108428.

Legendre, P., and L. Legendre. 1998. Numerical ecology. 2nd edition. Elsevier, Amsterdam, The Netherlands.

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MJM Publishers, Gleneden Beach, Oregon.

McKinley, D. C., A. J. Miller-Rushing, H. L. Ballard, R. Bonney, H. Brown, S. C. Cook-Patton, D. M. Evans, R. A. French, J. K. Parrish, T. B. Phillips, and S. F. Ryan. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. Biological Conservation 208:15–28.

Miller, D. L., R. M. Hughes, J. R. Karr, P. M. Leonard, P. B. Moyle, L. H. Schrader, B. A. Thompson, R. A. Daniels, K. D. Fausch, G. A. Fitzhugh, and J. R. Gammon. 1988. Regional applications of an index of biotic integrity for use in water resource management. Fisheries 13:12–20.

Moffett, E. R., and M. W. Neale. 2015. Volunteer and professional macroinvertebrate monitoring provide concordant assessments of stream health. New Zealand Journal of Marine and Freshwater Research 49:366–375.

Morley, S. A., and J. R. Karr. 2002. Assessing and restoring the health of urban streams in the Puget Sound Basin. Conservation Biology 16:1498–1509.

Naiman, R. J., and R. E. Bilby. 1998. River ecology and management in the Pacific coastal ecoregion. Pages 1–10 *in* R. J. Naiman and R. E. Bilby (editors). River ecology and management: Lessons from the Pacific Coastal Ecoregion. Springer, Berlin, Germany.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2018. Learning through citizen science: Enhancing opportunities by design. The National Academies Press, Washington, DC. (Available from: https://doi.org/10.17226/25183)

Nerbonne, J. F., and B. Vondracek. 2003. Volunteer macroinvertebrate monitoring: Assessing training needs through examining error and bias in untrained volunteers. Journal of the North American Benthological Society 22:152–163.

Nerbonne, J. F., B. Ward, A. Ollila, M. Williams, and B. Vondracek. 2008. Effect of sampling protocol and volunteer bias when sampling for macroinvertebrates. Journal of the North American Benthological Society 27:640–646.

O'Leary, N., A. T. Vawter, L. P. Wagenet, and M. Pfeffer. 2004. Assessing water quality using two taxonomic levels of benthic macroinvertebrate analysis: Implications for volunteer monitors. Journal of Freshwater Ecology 19:581–586.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2020. *vegan*: Community ecology package. (Available from: https://cran.r-project.org/web/packages/vegan/index.html)

Onion, A. M., A. J. Smith, and B.T. Duffy. 2023. New York State's WAVE Method-Evaluation of a method for water quality monitoring by citizen scientists using benthic macroinvertebrates. Environmental Management 71:432–438.

Overdevest, C., C. H. Orr, and K. Stepenuck. 2004. Volunteer stream monitoring and local participation in natural resource issues. Human Ecology Review 11:177–185.

Pandya, R. E. 2012. A framework for engaging diverse communities in citizen science in the US. Frontiers in Ecology and the Environment 10:314–317.

Penrose, D., and S. M. Call. 1995. Volunteer monitoring of benthic macroinvertebrates: Regulatory biologists' perspectives. Journal of the North American Benthological Society 14:203–209.

Peres-Neto, P. R., and D. A. Jackson. 2001. How well do multivariate datasets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129: 169–178.

Phillips, T. B., H. L. Ballard, B. V. Lewenstein, and R. Bonney. 2019. Engagement in science through citizen science: Moving beyond data collection. Science Education 103:665–690.

Pinto, P., J. M. B. Oliveira Jr, F. Leitão, M. M. Morais, L. Chícharo, P. Vaz, S. M. Delgado, C. Voreadou, E. A. Morales, and M. A. Teodósio. 2020. Development of a metric of aquatic invertebrates for volunteers (MAIV): A simple and friendly biotic metric to assess ecological quality of streams. Water 12:654.

Resh, V. H. 2008. Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. Environmental Monitoring and Assessment 138: 131–138.

Rieman, B. E., C. L. Smith, R. J. Naiman, G. T. Ruggerone, C. C. Wood, N. Huntly, E. N. Merrill, J. R. Alldredge, P. A. Bisson, J. Congleton, and K. D. Fausch. 2015. A comprehensive approach for habitat restoration in the Columbia Basin. Fisheries 40:124–135.

Rosenberg, D. M., and V. H. Resh. 1993. Freshwater biomonitoring and benthic invertebrates. Chapman Hall, New York, New York.

Silvertown, J. 2009. A new dawn for citizen science. Trends in Ecology & Evolution 24:467–471.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 4:878–891.

Thornhill, I., S. Loiselle, W. Clymans, and C. G. E. Van Noordwijk. 2019. How citizen scientists can enrich freshwater science as contributors, collaborators, and co-creators. Freshwater Science 38:231–235.

Tsui, L. 2007. Effective strategies to increase diversity in STEM fields: A review of the research literature. The Journal of Negro Education 76:555–581.

Valle, N., P. Antonenko, L. Endara, E. C. Davis, G. Somarriba, E. Sessa, F. Luo, S. Carey, S. Dogan, J. G. Burleigh, and S. McDaniel. 2021. Community science, storytelling, or inquiry-based learning? Evaluating three technology-enhanced pedagogical approaches in an online botany course. The American Biology Teacher 83:513–520.

Van Sickle, J., J. Baker, A. Herlihy, P. Bayley, S. Gregory, P. Haggerty, L. Ashkenas, and J. Li. 2004. Projecting the biological condition of streams under alternative scenarios of human land use. Ecological Applications 14:368–380.

Waite, I. R., L. R. Brown, J. G. Kennen, J. T. May, T. F. Cuffney, J. L. Orlando, and K. A. Jones. 2010. Comparison of watershed disturbance predictive models for stream benthic macroinvertebrates for three distinct ecoregions in western US. Ecological Indicators 10:1125–1136.

Walk, S. H. 1997. Volunteer stream monitoring: A methods manual. EPA 841-B-97-003. Office of Water, United States Environmental Protection Agency, Washington, DC. (Available from: https://archive.epa.gov/water/archive/web/html/stream_index.html)

Walker, D. W., M. Smigaj, and M. Tani. 2021. The benefits and negative impacts of citizen science applications to water as experienced by participants and communities. Wiley Interdisciplinary Reviews: Water 2021:e1488.

White, S. M., S. Brandy, C. Justice, K. A. Morinaga, L. Naylor, J. Ruzycki, E. R. Sedell, J. Steele, A. Towne, J. G. Webster, and I. Wilson. 2021. Progress towards a comprehensive approach for habitat restoration in the Columbia Basin: Case study in the Grande Ronde River. Fisheries 46:229–243.

Wilson, T. S., and D. G. Sorenson. 2012. Willamette Valley ecoregion. Pages 51–57 *in* B. M. Sleeter, T. S. Wilson, and W. Acevedo (editors). Status and trends of land change in the Western United States—1973 to 2000. United States Geological Survey Professional Paper 1794–A. (Available from: https://pubs.usgs.gov/pp/1794/a/)