SUPPORTING STATEMENT
NOAA BAY WATERSHED EDUCATION AND TRAINING (B-WET) PROGRAM
NATIONAL EVALUATION SYSTEM
OMB CONTROL NO. 0648-0658

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local governmental units, households, or persons) in the universe and the corresponding sample are to be provided in tabular form. The tabulation must also include expected response rates for the collection as a whole. If the collection has been conducted before, provide the actual response rate achieved.

Censuses will be conducted in light of the relatively small sample sizes (Error: Reference source not found) and the sophisticated analyses planned to be conducted by an external evaluator. More specifically, Stata data analysis and statistical software will be used to conduct confirmatory factor analysis, multilevel analysis (i.e., to account for teachers nested in professional development programs and repeated measures from the same individuals), and structural equation modeling (SEM) to explore the direct and indirect relationships between teachers' practices and perceived student outcomes based on their MWEE professional development experiences and background. Benefits of SEM include that it allows for exploring direct and indirect causal relationships between variables while also taking into account measurement error (Bollen, 1989). SEM permits the combination of factor and path analysis into a single model. SEM models require large sample sizes because they estimate: 1) regression coefficients, 2) variances and covariances of unobserved variables, and 3) variances and covariances of errors. Because of the number of direct and indirect paths that the models will estimate, they will have few degrees of freedom (df). These more sophisticated analysis will be conducted once the sample size is approximately 1,280, Based on an expected df=4 and the proposed sample size, a power of 80% will be achieved for testing model fit (see Table 4 in MacCallum, et al. 1996).

The expected response rates reported in Table 1 are informed by the response rates obtained since the initial implementation of the evaluation system in January 2014.

Table 1: Past and Expected Response Rates

Questionnaire	Time Period	N (number of emails sent)	Nᵃ (adjusted number of emails sent)	n (number who responded)	R (adjusted response rate)	Future expected R
Grantee	January 2014-June 2015	106	95	84	88%	90%
Teacher Post-PD	April 2014-July	436[b]	392	125	32%	40%

	2015					
Teacher Post-PD Nonresponse	July 2014 - July 2015		267[c]	44	16%	20%
Teacher Post-MWEE	March 2014-June 2015	546	491	163	33%	40%
Teacher Post-MWEE Nonresponse	June - July 2015		328[c]	50	15%	20%

[a]During the pilot-testing process in 2014, we discovered that about 20% of email invitations were not received by grantees, despite accurate contact information. Emails were likely rejected by respondents' servers. We tested the system again in 2014 with greater success. Qualtrics' email delivery success is measured by SenderReport.org. On 9/18/15, Qualtrics' average score was 87 out of a maximum score of 100 based on a review of the last 100 high volume emails sent out from the qemailserver.com server (a score of 70 or above is considered good). Although Qualtrics cannot provide a percent of emails that are rejected by servers, we believe 10% is an informed estimate. Thus, the estimated adjusted response rates (R) take into account the 10% of respondents who are potentially not receiving requests to complete the questionnaire.
[b]Qualtrics sent 493 emails to teachers, however 35 responded that they did not participate in B-WET PD, and 22 said they are not teachers; those 57 were subtracted from the N for teacher post-PD emails.
[c]Adjusted N minus respondents (i.e., 392-125=267 Post-PD Teachers; 491-163=328 Post-MWEE Teachers)

2. Describe the procedures for the collection, including: the statistical methodology for stratification and sample selection; the estimation procedure; the degree of accuracy needed for the purpose described in the justification; any unusual problems requiring specialized sampling procedures; and any use of periodic (less frequent than annual) data collection cycles to reduce burden.

Censuses of the respective populations will be conducted to attain the sample sizes needed for the proposed sophisticated analyses, which will allow for more in-depth answers to the evaluation system's questions.

3. Describe the methods used to maximize response rates and to deal with nonresponse. The accuracy and reliability of the information collected must be shown to be adequate for the intended uses. For collections based on sampling, a special justification must be provided if they will not yield "reliable" data that can be generalized to the universe studied.

Methods to Maximize Response Rates

Grantee Questionnaire
Since grantee data collection was initiated in January 2014, an 88% response rate has been achieved. This grantee response rate occurred as a result of (1) including information about the national evaluation in the B-WET federal funding opportunity (FFO), (2) providing preview copies of the evaluation system questionnaire on the B-WET website, (3) sending a pre-notification to all grantees at the beginning of their grant year (Dillman et al., 2009), and (4) sending two reminder invitations, two and four weeks following the initial invitation (Dillman et al., 2009), to complete the questionnaire at the end of their grant year. Because B-WET grantees receive funds from NOAA to conduct their MWEE projects, they are highly invested in the B-WET program. In light of these practices and circumstances, as well as the 88% grantee response rate achieved by the evaluation system to date (Table 1), a 90% grantee response rate is therefore

expected in the future.

Teacher Post-PD Questionnaire
Since teacher post-PD data collection was initiated in April 2014, a 32% response rate has been achieved. This is in contrast to the initially proposed 80% response rate which would have been achievable if this questionnaire could have been administered as a final activity of the teacher professional development (Zint, 2010, 2009, 2008). Based on what has been learned from developing and pilot-testing the evaluation system, however, it is not possible to administer this questionnaire as part of the professional development using the on-line evaluation system. For one, teachers need to respond to individualized links with embedded data (matched with respective grantee information to support planned analyses) and second, it would also not be possible to send targeted reminders to complete the questionnaire to non-respondents (Dillman et al., 2009; Yu and Cooper, 1983).

In light of the experience so far, a 40% response rate is expected in the future. This response rate will be achieved (1) by following Dillman et al.'s (2009) recommendations for survey design and administration (e.g., advance notice about questionnaire; multiple, personal requests to complete questionnaire; use of closed-ended, clear, easy to understand questions; etc.), (2) as a result of grantees' greater familiarity with the data collection process, as well as (3) by continuing and improving current activities to raise teacher response rates by the national coordinator, regional coordinators, and grantees. For example, the national coordinator will continue and enhance national evaluation system webinars for grantees, raise awareness of the evaluation resources available through the B-WET evaluation website, send monthly reminders to grantees to add teacher emails to the evaluation system, and participate in meetings with grantees to discuss ways to increase teacher participation in the national evaluation. The regional coordinators will further refine content about the national evaluation as part of their FFOs (e.g., by asking grantees to describe how they will participate in the national evaluation system) and play a much greater role than they have so far in encouraging grantees to encourage their teachers' participation in the evaluation (e.g., meetings about the national evaluation with the grantees that may include the national coordinator). In addition, a question has been added to the grantee questionnaire to learn how they are encouraging teachers to participate in the national evaluation so that this information can be used to further improve the process for maximizing teacher response rates.

Teacher Post-MWEE Questionnaire
Since teacher post-MWEE data collection was initiated in March 2014, a 33% response rate has been achieved, lower than the 40% initial expected response rate. This 40% response rate was expected based on: (1) similar evaluations (i.e., ones including Internet-based questionnaires administered within same time frames after teacher professional development by the providers of these programs) offered by environmental educators that have resulted in 35-80% response rates (Zint, 2010, 2009, 2008) and (2) following Dillman et al.'s (2009) recommendations to maximize response rates. These recommendations include (a) asking grantees to inform their teachers in advance that they will be asked to complete this questionnaire as part of their professional development responsibilities, (b) use of multiple, personalized completion requests by grantees with whom teachers are familiar, (c) the use of questions that are closed-ended, worded in a clear, easy to understand manner, (d) the use of skip logic, and (e) asking for little personal or sensitive information (Dillman et al., 2009; Dillman, Sinclair, and Clark, 1993). The future expected response rate is once again 40% and will be achieved based on further refinements to the above process as well as based on grantees' greater familiarity with the data

collection process.

Nonresponse Analysis
The evaluation system is designed to include a post-PD and post-MWEE nonresponse questionnaire, to ensure that comparisons can be made between initial and non-respondent teachers, when response rates are below 80%. In this instance, B-WET has, and will in the future, engage an external contractor to conduct analyses of these results.

All non-respondents receive an email invitation with a Web link to an abbreviated version of the questionnaires (automatic reminders are sent once). Results from these questionnaires have been, and will in the future be compared with those from earlier respondents to determine if there are statistically and substantively significant differences.

If earlier respondent and non-respondent populations are determined not to be significantly and substantively different, no further analysis will occur. If it is determined that the non-respondent population is significantly and substantively different from the earlier respondent population, analysis with weighted adjustments for nonresponse, using a method such as those described in Part IV of *Survey Nonresponse* (Groves et. al. 2002), will be conducted.

Teacher Post-PD Nonresponse Questionnaire
Although the response rate for this questionnaire has been lower than initially expected (r=32% vs. r=80%), based on the shorter nonresponse post-PD questionnaire, there is evidence to suggest that the data collected to date can be generalized to the population of NOAA B-WET participating teachers. Only 2 of 30 statistical significant tests indicated that there were statistically significant differences between initial and non-respondent teachers. Importantly, there were no statistically significant differences in the programs teachers experienced nor in outcomes. The only differences occurred in the subjects teachers taught (e.g., nonresponse teachers were more likely to be science and math teachers whereas initial respondents were more likely to teach multiple disciplines including science) and that non-respondents were slightly more confident in their ability to implement MWEEs before their professional development than initial respondents (Attachment 3). At the same time, it is acknowledged that the two sample sizes (initial questionnaire n=133, non-respondent questionnaire=23) were not large enough to have sufficient power (Cohen, 1988) to detect statistically significant differences between the two groups (Attachment 4). Nonetheless, the effect sizes corresponding to the differences ranged from .032 to .362 (mean=.17) for continuous variables (where means were compared) and from 0.00 to .08 (mean=.04) for binary variables (where proportions were compared), meaning that the effect sizes were "small" per Cohen (1988). This suggests that the differences between initial and non-respondents were in fact small and our inability to detect the differences as significant were not just due to our sample sizes. In addition, a qualitative inspection of the two groups' responses to the questionnaire's common measures shows no substantively meaningful differences between the two groups' means and frequency responses (Attachment 3).

Teacher Post-MWEE Nonresponse Questionnaire
Based on the 33% response rate achieved for the post-MWEE questionnaire so far, an analysis was conducted comparing results from the initial post-MWEE questionnaire with the much shorter nonresponse questionnaire. Findings suggest that data collected through the post-MWEE questionnaire to date can be generalized to the population of NOAA B-WET participating teachers. Only 5 of 22 statistical significant tests indicated that there were statistically significant

differences between initial and non-respondent teachers (Attachment 5). Two of the differences were in the implementation of MWEEs with non-respondent teachers reporting that they conducted slightly shorter MWEEs and that their students spent slightly less time outdoors than initial respondents. Three of the differences occurred in the perceived student outcomes of MWEEs (12 in total), with non-respondent teachers reporting slightly less confidence that their students achieved the respective outcomes. At the same time, it is acknowledged that the two sample sizes (initial questionnaire n=113, non-respondent questionnaire=23) were not large enough to have sufficient power (Cohen, 1988) to detect statistically significant differences between the two groups (Attachment 4). Nonetheless, the effect sizes corresponding to the differences ranged from .01 to .41 (mean=.17) for continuous variables (where means were compared) and were .11 for each of the two binary variables (where proportions were compared), meaning that the effect sizes were "small" per Cohen (1988). This suggests that the differences between initial and non-respondents were in fact small and our inability to detect the differences as significant were not just due to our sample sizes. In addition, a qualitative inspection of the two groups' responses to the questionnaire's common measures shows no substantively meaningful differences between the two groups' and means and frequency responses (Attachment 5).

4. Describe any tests of procedures or methods to be undertaken. Tests are encouraged as effective means to refine collections, but if ten or more test respondents are involved OMB must give prior approval.

The majority of measures and procedures used as part of the B-WET evaluation system have been tested and successfully implemented by previous studies (e.g., "Evaluation of National Oceanic and Atmospheric Administration Chesapeake Bay Watershed Education and Training Program," Kraemer et al., 2007; Zint et al., 2014). In addition, an exploratory study of the benefits of MWEEs found that the scales used as part of the proposed B-WET evaluation system (examined using exploratory factor analysis in SPSS and M+) are reliable and valid (Zint, 2012). Reliabilities, for example, ranged between good and excellent (i.e., Cronbach Alpha range: .70 to .90) and the amount of variance explained by the factors were substantial (i.e., range: 40% to 90). The measures used as part of the evaluation system have also been examined for face and content validity by stakeholders consisting of the nine members of NOAA's internal B-WET Advisory group, three evaluation experts with knowledge of B-WET, three B-WET grantees, and two watershed scientists.

As part of this application, some revisions to the three questionnaires are being requested, based on a review of descriptive statistics of initial data as well as respondents' feedback.

No additional testing is planned.

5. Provide the name and telephone number of individuals consulted on the statistical aspects of the design, and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Individuals Consulted on Statistical Design:
Dr. Michaela Zint, Professor, School of Natural Resources & Environment, School of Education, and College of Literature, Science & the Arts at the University of Michigan developed the statistical design for the proposed evaluation system. She, in turn, consulted with:

- Dr. Heeringa & Statistical Design Group members, Institute for Social Research, University of Michigan
- Dr. Lee & Dr. Rowan, School of Education, University of Michigan
- Dr. Rutherford & Dr. West, Center for Statistical Consultation and Research, University of Michigan

If you have any questions about the statistics design of the study, please contact Dr. Michaela Zint: zintmich@umich.edu, 734.763.6961.

Individual Who Will Conduct Data Collection and Analysis:
The evaluation system is designed to collect data through Qualtrics, an online survey platform that automatically generates descriptive statistics. Data may also be downloaded from Qualtrics for more sophisticated analysis by an external contractor.

Bronwen Rice, B-WET National Coordinator, NOAA Office of Education (Bronwen.Rice@noaa.gov, 202.482.6797) will be responsible for managing the data collection process and for ensuring the functioning and maintenance of the evaluation system.

LITERATURE CITED

Bollen, K.A. 1989. Structural equations with latent variables. New York: Wiley.

Burton, L.J. and Mazerolle, S.M. 2011. Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research. Athletic Training Education Journal. Vol. 6, No. 1, 27-35.

Carmines, E. G. and Zeller, R. A. 1979. Reliability and Validity Assessment. Sage: Beverly Hills, CA.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences (2nd ed). Lawrence Erlbaum Associates: Hillsdale, N.J., p. 567.

Dillman, D.A., Sinclair, M.D., and Clark, J.R. 1993. Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. Public Opinion Quarterly. Vol. 57, 289-304.

Dillman, D. A., Smyth, J.D. and Christian, L.M. 2009. Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method, 3rd edition. John Wiley: Hoboken, NJ.

Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A. 2002. Survey Nonresponse. John Wiley & Sons, Inc.: New York.

Kraemer, A., Zint, M., and Kirwan, J. 2007. An Evaluation of National Oceanic and Atmospheric Administration Chesapeake Bay Watershed Education and Training Program Meaningful Watershed Educational Experiences. Unpublished. http://chesapeakebay.noaa.gov/images/stories/pdf/Full_Report_NOAA_Chesapeake_B-WET_Evaluation.pdf

Litwin, Mark S. 1995. How to Measure Survey Reliability and Validity. Sage: Thousand Oaks, CA.

MacCallum, R. C., Browne, M. W, and Sugawara, H. M. 1996. Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, Vol. 1, No. 2, 130-149.

Nunally, J. C. and Bernstain, I.H. 1994. Psychometric Theory. McGraw-Hill: New York.

Patton, M.Q. 2008. Utilization-focused evaluation. 4th ed. Sage: Los Angeles.

Qualtrics. 2013. Qualtrics Security White Paper: Why should I trust Qualtrics with my sensitive data? https://www.utexas.edu/its/downloads/survey/2185/White%20Paper_Qualtrics

%20Security_1%2018%2013%20(2).pdf

U. S. Department of Labor, Bureau of Labor Statistics. May 2011. National Compensation Survey: Occupational Earnings in the United States, 2010. Table 5: Full-time State and local government workers: Mean and median hourly, weekly, and annual earnings and mean weekly and annual hours: http://www.bls.gov/ncs/ocs/sp/nctb1479.pdf.

Yu, J. and Cooper, H. 1983. A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. Journal of Marketing Research. Vol. XX, 36-44.

Zint, M. 2008, 2009 & 2010. Summary of annual Environmental Education and Training Partnership achievements. Annual reports to the U.S. Environmental Protection Agency's Office of Environmental Education. University of Michigan: Ann Arbor, MI.

Zint, M. 2011. A literature review of watershed education-related research to inform NOAA B-WET's evaluation system. University of Michigan: Ann Arbor, MI.

Zint, M. 2012. An exploratory assessment of the benefits of MWEEs: A report prepared for NOAA. University of Michigan: Ann Arbor, MI.

Zint, M., Kraemer, A. & Kolenic, G. E. 2014. Evaluating Meaningful Watershed Educational Experiences: An exploration into the effects on participating students' environmental stewardship characteristics and the relationships between these predictors of environmentally responsible behavior. Studies in Educational Evaluation: Special Issue on Research in Environmental Education Evaluation, Vol. 41, 4-17.

ATTACHMENTS (posted as supplementary documents)

1.   Evaluation System Metrics Matrix
2a-e.   Revised Questionnaires: Grantee, Teacher Post-PD, Teacher Post-PD Nonresponse, Teacher Post-MWEE, Teacher Post-MWEE Nonresponse
3.   Comparison of Teacher Post-PD Initial vs. Nonresponse Questionnaire Results
4.   Teacher Post-PD and Post-MWEE Power and Effect Size Analysis Results
5.   Comparison of Teacher Post-MWEE Initial vs. Nonresponse Questionnaire Results