**To: Dr. Ashley Shade**

**From: STEM E.D., LLC (Libarkin)**

**Re: Edamame Workshop Evaluation**

**Date: Sept 10, 2014. FINAL.**

A pre-workshop evaluation of the Edamame Workshop 2014 was conducted on August 13, with a post-workshop evaluation occurring August 20, 2014. Observations were also conducted at the start and end of the workshop. In all, 23 participants completed the pre-survey and 21 completed the post-survey; 20 participants completed both surveys.

***Demographics***

Demographic data are reported for pre-workshop respondents. One participant declined to respond to the age question. Remaining participants ranged in age from 23-43, with an average age of 27.9 ± 5.3 years. Participants were 52% male (n=12), with the remainder female (n=11) and no transgendered participants. Participants were also 61% exclusively Caucasian (n=14). Remaining participants were: Latino(a) (n=3); African/African American/Black (n=2); and Caucasian/White and Native American (n=1). Two participants declined to respond to the race/ethnicity question. Finally, the academic status of participants included graduate students (n=15) and Ph.D.-holding professionals (n=8).

**EXECUTIVE SUMMARY OF RESULTS**

We summarize evaluation results below. Following this summary, we provide: 1) A description of scales used to measure participant characteristics, including indices of validity, reliability, and overall scores, and 2) results for qualitative and close-ended questions related to participant impressions of the workshop. In summary, we found that:

1. Observations suggest that *the structure of the EDAMAME workshop and the built-in methods for interaction between leaders and students were excellent.*
2. Scores on the Perception of Computational Ability scale were calculated for both the pre- and post-workshop surveys. Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different (Z = -3.933, p ≤ 0.001), with higher post-workshop scores. *This indicates that participants perceived greater computational ability after engagement in the workshop.*
3. Scores on the Computational Understanding scale were calculated for both the pre- and post-workshop surveys. Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different (Z = -3.925, p ≤ 0.001), with higher post-workshop scores. *This indicates that participants perceived greater computational understanding after engagement in the workshop.*
4. Scores on the Coding Ability scale were calculated for both the pre- and post-workshop surveys. Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different (Z = -3.922, p ≤ 0.001), with higher post-workshop scores. *This indicates that participants perceived greater coding ability after engagement in the workshop.*
5. *Participants were generally very satisfied with the workshop. On average, participants rated the workshop components as Good-Very Good, where Very Good was the maximum score.*
6. *Participants felt the workshop met their needs and would overwhelmingly recommend it to others.*
7. Participants were generally positive about the workshop in their open-ended comments. Suggestions for improvement include: *stick to scheduled breaks, offer more explanation of the workshop plans ahead of time, and lengthen the workshop to two weeks.*

**WORKSHOP EVALUATION AND RESULTS**

The workshop was evaluated through observation, Likert-type scales, and open-ended prompts. Scales used for the workshop were modified from scales used in earlier evaluation of similar workshops (Software Carpentry; Bioinformatics); scale statistics were calculated to ensure validity and reliability of scales for this workshop. Participants completed identical pre- and post-workshop surveys containing three scales: Perceptions of Computational Ability, Computational Understanding, and Coding Ability. Respondents also completed several demographic questions and responded to two open-ended questions and one box-and-arrow prompt covering workshop content. The surveys also contained questions about expectations (pre) and overall perceptions (post) of the workshop. Scale analyses were performed on pre-instruction results except where noted.

***Perception of Computational Ability***

Participants rated their ability in each as No Ability, Low Ability, Intermediate Ability, or High Ability on nine questions related to computational ability. A score of 1 implies No Ability, while a score of 4 implies High Ability. One scale measuring perception of computational ability in general was identified and consisted of five of these nine items. An average across these nine items is used to calculate a score for Perception of Computational Ability.

**Table 1. Factor loadings, communalities, and item response averages for Perception of Computational Ability Scale.**

|  |  |  |
| --- | --- | --- |
| **Items** | **Factor Loadings** | **Communalities** |
| Navigating terminal | .617 | .381 |
| Executing Python scripts | .724 | .524 |
| Using Amazon EC2 | .773 | .597 |
| Installing and running mother | .639 | .409 |
| Executing R scripts | .731 | .534 |

*Factor Analysis.* Exploratory factor analysis of pre-instruction responses to Perceptions of Computational Ability items was undertaken to investigate the presence of one or more scales. Of the nine items, five satisfactorily constituted a single factor, while the remaining four items were inconsistently correlated. The five-item scale was explored.

Note that this analysis is preliminary given the small sample size, although the Kaiser-Meyer-Olkin measure of sampling adequacy for the five items was 0.56, close to the 0.6 value recommended for factor analysis. The data set also met other minimum conditions necessary for factor analysis. First, items correlated with at least one other item at a level over 0.3, indicating that a factor structure could be expected to emerge. The Bartlett’s Test of sphericity was significant (χ2 (10) = 33.48, *p*<0.001). Communalities for all items were above 0.3, indicating shared variance with other items. A single scale explains 48.9% of the variance in the data and all items yielded factor loadings well over 0.32, suggesting presence of a single scale (Table 1).

*Reliability analysis.*Cronbach’s alpha was calculated to establish internal consistency of items. The scale has high reliability, with a calculated alpha for the sample of 0.729. Despite the small sample, this is above the recommended minimum of 0.70 for analysis of individual scores.

*Results.* Scores on the Perception of Computational Ability scale were calculated for the pre- and post-workshop survey. Overall scale scores were calculated as averages across all five items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data (n=20). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different (Z = -3.933, p ≤ 0.001). Statistical significance and mean results (Table 2) indicate that perceptions of computational ability increased as a result of instruction, from Low/No Ability to Intermediate/High Ability.

**Table 2. Pre and Post Workshop Scale Averages**

|  |  |  |
| --- | --- | --- |
| **Scale** | **Pre** | **Post** |
| Perception of Computational Ability | 1.97 ± 0.56  (No to Low Ability) | 3.18 ± 0.32  (Intermediate to High Ability) |
| Computational Understanding | 2.25 ± 0.47  (Disagree to Agree) | 3.37 ± 0.33  (Agree to Strongly Agree) |
| Coding Ability | 2.30 ± 0.44  (Disagree to Agree) | 3.32 ± 0.28  (Agree to Strongly Agree) |

***Computational Understanding***

Participants were asked to indicate the extent to which they understood specific concepts, with ratings of Strongly Disagree, Disagree, Agree, and Strongly Agree. A score of 1 implies low understanding (Strong Disagreement); a score of 4 implies high understanding (Strong Agreement).

*Factor Analysis.* Exploratory factor analysis of responses to pre-workshop Conceptual Understanding items was undertaken to evaluate the presence of one or more scales. Five factors emerged based on eigenvalue analysis, while one strong factor was suggested by scree plot analysis. Based on this, confirmatory analysis for one scale was conducted. This analysis is preliminary given the small sample size, with a Kaiser-Meyer-Olkin measure of sampling adequacy of 0.58, just below the 0.6 value recommended for factor analysis. The data set met other minimum conditions necessary for factor analysis. First, the majority of items correlated with at least one other item at a level over 0.3, indicating that a factor structure could be expected to emerge. The Bartlett’s Test of sphericity was significant (χ2 (78) = 168.1, *p*<0.001). Communalities for many items were above 0.3, indicating shared variance with other items. Given these data, confirmatory factor analysis was performed on all sixteen items. Subsequent removal of three items for poor loading yielded one satisfactory scale. This single scale explains 40.3% of the variance in the data and all items yielded factor loadings over 0.32, again suggesting the presence of a single scale (Table 3).

**Table 3. Factor loadings and communalities for Computational Understanding Scale**

|  |  |  |
| --- | --- | --- |
| **Items** | **Factor Loadings** | **Communalities** |
| I know how to process Illumina data | .364 | .132 |
| I understand whatper\_library\_statspy does | .502 | .252 |
| I know how to run R | .450 | .203 |
| I know the main differences in analyses offered by QIIME and mot | .785 | .616 |
| I am familiar with biom formatted files | .639 | .408 |
| I can name at least two different microbial metagenomic databases | .842 | .709 |
| I know what an R package is | .485 | .236 |
| I understand the structure of an OUT table | .735 | .540 |
| I know what a kmer is | .629 | .396 |
| I know the difference between alpha and beta diversity | .639 | .409 |
| I know how to visualize microbial metagenomic data | .686 | .471 |
| I know how to use metadata to guide community analyses | .847 | .717 |
| I know how to assemble shotgun metagenomic data | .388 | .151 |
| I know how to process Illumina data | .364 | .132 |

*Reliability analysis.*Cronbach’s alpha was calculated to establish internal consistency of items. The scale has high reliability, with a calculated alpha for the sample of 0.868. Despite the small sample, this is above the recommended minimum of 0.70 for analysis of individual scores.

*Results.* Scores on the Computational Understanding scale were calculated for both the pre- and post-workshop surveys. Overall scale scores were calculated as averages across all 13 items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data (n=20). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different (Z = -3.925, p ≤ 0.001). Statistical significance and mean results (Table 2) indicate that computational understanding increased as a result of instruction.

***Coding Ability***

A scale measuring coding ability was modified from Python-related scales used in prior evaluations, itself modified from Askar and Davenport (2009), to focus on concepts taught in the EDAMAME workshop. A set of 27 items were developed. Participants were asked to indicate the extent to which they agreed with statements about their coding ability, with ratings of Strongly Disagree, Disagree, Agree, and Strongly Agree. A score of 1 implies low ability (Strong Disagreement), while a score of 4 implies high ability (Strong Agreement). Inversely worded items, those containing the word CANNOT, were reverse coded before reliability analysis.

*Factor Analysis.* Iterative removal of poorly loading items resulted in identification of one scale consisting of 17 items. Scree plot and eigenvalue analysis suggests the presence of one dominant scale, prompting consideration of the items as a single scale. Note that this analysis is preliminary given the small sample size, with a Kaiser-Meyer-Olkin measure of sampling adequacy of 0.501, below the 0.6 value recommended for factor analysis. The data set met other minimum conditions necessary for factor analysis. First, the majority of items correlated with at least one other item at a level over 0.3, indicating that a factor structure could be expected to emerge. The Bartlett’s Test of sphericity was significant (χ2 (136) = 279.6, *p*<0.001). Communalities for most items were above 0.3, indicating shared variance with other items.

A single scale explains 37.1% of the variance in the data and all items yielded factor loadings over 0.32, suggesting the presence of a single scale (Table 4).

**Table 4. Factor loadings and communalities for Coding Ability Scale**

|  |  |  |
| --- | --- | --- |
| **Items** | **Factor Loadings** | **Communalities** |
| I can independently execute the steps of microbial AMPLICON metagenomic analysis from raw data processing to statistics. | .532 | .283 |
| I can make informed choices for analysis parameters (e.g., OTU clustering, tree building, denoising) for AMPLICON analyses in mothur. | .434 | .189 |
| I understand the differences between unassembled and assembled shotgun metagenomic data. | .647 | .419 |
| I understand the limitations of unassembled and assembled shotgun metagenomic data. | .462 | .213 |
| I understand what digital normalization is. | .564 | .318 |
| I understand why I might use digital normalization. | .573 | .328 |
| I understand what a k-mer is. | .821 | .674 |
| I understand what k-mer means in the context of assembly. | .774 | .600 |
| I can use an R package to analyze data. | .469 | .220 |
| I can move data out of mothur and into R. | .710 | .505 |
| I can find and move files on a computer using terminal. | .626 | .391 |
| I can evaluate the quality of raw sequencing data. | .552 | .305 |
| I can deposit a sequencing dataset and make it publicly accessible. | .693 | .480 |
| I can search for public datasets that are relevant to my own project. | .615 | .379 |
| I CANNOT process and quality control AMPLICON metagenome data unless someone else helps me get started. | -.672 | .451 |
| I can find ways of overcoming the problem if I get stuck at a point during data analysis. | .605 | .366 |
| I can design an experiment collecting 16S marker gene data that someone else can comprehend. | .435 | .189 |

*Reliability analysis.*Cronbach’s alpha was calculated to establish internal consistency of items. The scale has a very high reliability, with a calculated alpha for the sample of 0.886. Despite the small sample, this is above the recommended minimum of 0.70 for analysis of individual scores. Note that this value was calculated after reverse coding of four negatively loaded items, as is appropriate for the alpha metric.

*Results.* Scores on the Coding Ability scale were calculated for both the pre- and post-workshop surveys. Overall scale scores were calculated as averages across all 17 items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data (n=20). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different (Z = -3.922, p ≤ 0.001). Statistical significance and mean results (Table 2) indicate that coding ability increased as a result of instruction.

***Open-Ended Questions and Box-and-Arrow Diagram***

Two open-ended questions and one box-and-arrow diagram related to the concepts covered in the workshop were completed pre- and post-workshop (Table 5). Open-ended questions were analyzed for evidence of increased understanding of workshop concepts, specifically as reflected in the complexity of participant responses. Response rates were higher post-workshop for individual questions, suggesting greater ability to respond to questions; as one participant put it pre-workshop, “*I am not familiar with how to analyze data using the Illumina platform.*” In general, complexity of responses increased from pre- to post-workshop, with much more discussion of what the participants did not know on pre-workshop surveys. Further analysis of open-ended responses for accuracy is possible with collaboration between evaluators and workshop facilitators. The box-and arrow diagram asked participants to “use the Illumina platform to evaluate both the phylogenetic diversity and functional potential” of a new microbial community. In general, pre-workshop diagrams were much simpler than post-workshop. Further analysis of box-and-arrow diagrams should occur in collaboration with the workshop facilitators.

**Table 5. Open-Ended Response Rates**

|  |  |  |
| --- | --- | --- |
| **Open-Ended** | **Pre Response Rate\*** | **Post Response Rate\*\*** |
| 1. Suppose that you are doing a 16S marker analysis of the gut microbial community from a non-model worm. This worm has a unique digestive tract evolved for the breakdown of a specialized diet. Discuss a sequencing strategy, a number of challenges you expect to face, and what kinds of additional resources are needed to help explore your data set. | 87% | 95% |
| Suppose that you are using the Illumina platform to sequence total DNA from soil samples. For this particular study, the soil samples come evenly distributed from two plot types on a single farm - one growing corn, the other dairy cow pasture. Discuss one or more approaches you would take to analyze the data, as well as your expected sensitivity and specificity to taxonomic and functional sequence identification. Include in your discussion the technical limitations of your analysis. | 70% | 90% |

\*Response rates are for participants completing surveys only; “no idea” or similar responses are not included in response rate. \*\*One participant indicated that they knew the answer, did not have enough time to explain for both post-survey questions. This is coded as a non-response.

***Observations***

The workshop was observed twice, once at the beginning and once at the end. The EDAMAME workshops were very structured and designed to encourage interaction among the students and instructor. Students appeared attentive and interested in the topics covered in the workshop. The workshop had an interesting tool in place (colored Post-It notes) that students used to signal the instructors if they needed additional help. The visuals used by speakers were acceptable, although there were times when the font size was too small or the audio was difficult to follow. Topic transition was somewhat disjointed. Overall, the structure of the EDAMAME workshop and the built-in methods for interaction between leaders and students were excellent. The workshop may want to set standards for visuals used by speakers and may want to develop new mechanisms for topic transition.

***Overall Workshop Impressions***

Participants were asked to respond to close-ended questions related to overall impressions of the workshop (Table 6). Participants also responded to open-ended questions about their expectations for the workshop (pre-survey) and their perceptions of the workshop (post-survey).

**Table 6. Overall Workshop Impressions**

|  |  |  |
| --- | --- | --- |
| **Workshop Components** | **Average Score**  **(ideal score)** | **% Yes** |
| 1. Concept: Alpha Diversity | 4.52 (5) | NA |
| 1. Concept: Beta Diversity | 4.57 (5) | NA |
| 1. Amplicon analyses in QIIME | 4.38 (5) | NA |
| Amazon EC2 | 3.90 (5) | NA |
| Amplicon analyses in mothur | 4.52 (5) | NA |
| Functional analyses with shotgun metagenomics | 3.19 (5) | NA |
| Shotgun metagenomics assembly | 3.05 (5) | NA |
| High Performance Computing Cluster (HPCC) | 4.15 (5) | NA |
| Overall Workshop Rating | 4.57(5) | NA |
| Overall Invited Speaker Rating | 4.81 (5) | NA |
| Meet Needs? | 3.67 (4) | NA |
| Learn What Hoped to Learn? | NA | 95% |
| Computational Understanding Change? | NA | 95% |
| Recommend Workshop? | NA | 100% |

*Close-Ended Questions.* Twelve questions asked participants to rate components of the workshop, as well as the overall workshop and speakers, on a 5-point Likert scale of Very Poor-Poor-Adequate-Good-Very Good. A 1 corresponds to a Very Poor rating and a 5 corresponds to a Very Good Rating. Participants were also asked if the workshop met their needs on a 4-point scale (4=very well), if they learned as expected from the workshop, if understanding of computational science changed, and if they would recommend the workshop.

*Results.* Post-workshop responses to questions about the efficacy of the workshop indicate that participants were generally very satisfied with the workshop (Table 6). In general, participants rated the workshop components as Good-Very Good. Two exceptions (Functional analyses with shotgun metagenomics; Shotgun metagenomics assembly) were rated Adequate-Good. Both concepts were related to shotgun metagenomics; subsequent comments on the survey suggest that this part of the workshop did not function correctly. Workshop leaders may want to reconsider how this content is taught.

Participants generally felt the workshop met their needs and would overwhelmingly recommend it to others. Participants gave the workshop a rating of 4.57 out of 5, indicating the workshop met their needs well to very well. Ninety-five percent (n=20) of participants felt their computational understanding changed and that they had learned what they hoped to learn, while 100% (n=21) would recommend the workshop to others.

*Qualitative Data: Participant Expectations and Perceptions.* On the pre-survey, participants responded to a prompt: “*Please provide any additional comments about your expectations for the workshop below*.” Participants also responded to a similar post-survey question: “*Please provide any additional suggestions or comments about the workshop below.*”, and were given opportunities to comment to each of the three yes-no questions in Table 6.

*Results.* Eleven participants provided comments about expectations on the pre-survey (Table 7). These comments related to interest in analyzing sequencing data, use of multiple data sets, and increasing understanding of tools. Given their brevity, pre-workshop expectations are provided verbatim:

* *I expect to learn how to take Illumina-type data and go through the process of data analysis.*
* *I am thrilled to be here!*
* *I hope to gain the necessary skill in order to successfully complete my graduate thesis project.*
* *Looking forward to learning how to utilize R for statistical analysis. Additionally, the metagenomic data analysis is very interesting. I only have experience using QIIME so everything else will be new.*
* *I expect to learn a lot about how to conduct metagenomic analyses.*
* *By the end I hope to know how to: 1) Perform a metagenomic analysis from DNA to Annotation; 2) Perform and determine applicable statistical analyses to compare microbial data; Use mothur*
* *I expect to leave this workshop with the knowledge and confidence to analyze large microbial datasets and to be able to properly answer all of these questions.*
* *I hope to understand the bioinformatic steps of analyzing sequence data more thoroughly, and at least obtain a basic understanding of the concepts. I would like to be able to know the process from A-Z so I can at least know what questions to ask and how to troubleshoot problems with my data appropriately.*
* *Can't wait to learn the answers to these questions!*
* *Its the first time I have next-gen sequencing data and I would like to have a better idea how to use the software to analyze my data.*
* *Expectations are very high since my academic formation was not related to metagenomic or bioinformatics. i expect this course will help me to clarify some concepts and be more confident to analyse my data*

Eighteen post-survey responses expressed either satisfaction with the workshop or suggestions for improvement. In general, comments were positive. Participants appreciated the invited speakers, the access to tutorials, and the availability of resources after the workshop. Examples are provided below:

* *Overall, this was a really great workshop and I learned a lot. The instructors were excellent in terms of their knowledge, explanations, and patience. Additionally, the resources provided are outstanding. All the the [sic] tutorials are well thought out, easy to access, and easy to understand.*
* *I cannot thank you all enough for putting on such an amazing workshop*
* *The invited speakers were great!*

Suggestions for the workshop included: Offering more breaks and/or sticking to scheduled breaks (n=4; note – one participant felt losing the afternoon off was difficult), making the workshop longer by lengthening it to two weeks (n=2), make sure EC2/shotgun data analyses would work before the workshop (n=2), increase time on assembly and annotation (n=1), break metagenomics into smaller parts (n=1), longer explanation of mother (n=1), designated office hours for help with data analysis (n=1), dedicated time for participant data analysis (n=1), better travel/pick-up information (n=1), and more time to complete the evaluation surveys (n=1). One participant suggested providing more details about daily activities, specifically: 1. *Having more of an overview (preferably visual) of all the programs (and steps within some programs) and where everything fits in. Then keep referring to flow charts and where we are in them. 2. Posting talks and tutorials online before they were presented. Those that had that option were definitely smoother.*

**Table 7. Perception and Expectations Response Rates**

|  |  |
| --- | --- |
| **Prompt/Question** | **N responses** |
| 1. Please provide any additional comments about your expectations for the workshop. (pre) | 7 |
| 1. Please provide any additional suggestions or comments about the workshop. (post) | 18 |

**REFERENCES**

Askar, P. & Davenport, D. (2009). An investigation of factors related to self-efficacy for Java programming among Engineering Students. Turkish Online Journal of Educational Technology, 8(1), 26-32.

END REPORT.