

Validation of the Amavita Interest Inventory: Executive Summary

Michelle Riconscente, PhD
Designs for Learning, Inc.

The goal of the present study was to validate the Amavita Interest Inventory, a reimagined assessment of Holland's (1997) six core dimensions of occupational interest based on the previously validated O*NET Interest Profiler Short Form. An additional purpose of the study was to determine whether or not respondents preferred the new version over the traditional one.

The validity study was carried out in two phases. The first phase was a pilot with half of the inventory items to confirm that the animated approach to item presentation held promise of validity and reliability. Following revisions to the piloted items and development of the remaining thirty items, a final validity study phase was carried out using the entire set of items on both the Amavita and O*NET inventories.

- Results from the pilot study of 225 respondents confirmed the potential of the Amavita visual approach to attain high consistency with the original O*NET version, and to do so in a more engaging and appealing format. Based on the pilot results, eight items were revised to increase comprehension; language prompts were added to all items; and an “unsure” option was added to the response scale.
- Main study participants (n=175) were 61% male, 74% Caucasian and ranged in age from 18-30 years (mean=27.2, sd=2.38). Respondents were distributed over 41 states and represented all twelve of the national regions defined by the Federal Reserve. Half (51%) held a bachelor's degree or higher.
- **Results from the main concurrent validity study suggest that the Amavita version of the O*NET Short Form Interest Profiler produces scores that are both reliable and valid.**
 - Scale correlations between the Amavita and O*NET versions ranged from .89 to .94, with all correlations significant at $p < .001$.
 - Reliability coefficients for all six Amavita scales demonstrated acceptable values, ranging from 0.80 to 0.93; reliability for O*NET Short Form scales ranged from 0.86 to 0.94.
 - Individuals' maximum scale scores on each measure were compared for consistency. The majority (87%) of participants' highest scores were on the same dimension on both versions.
- **Participants rated the Amavita version as more engaging relative to the traditional presentation format.**
 - Respondents rated Amavita as being statistically significantly more fun, more enjoyable, more entertaining, and less boring than the text version.
 - Three out of four respondents said they would choose the Amavita version over the text version of the inventory.
 - The majority (70%) reported liking the Amavita item format more than the text format.

Introduction

In today's employment landscape, a pressing challenge is matching job seekers to careers in which they will thrive and find satisfaction. Research has shown that the majority of United States workers are unhappy in their jobs¹, and increasingly employers are recognizing the importance of identifying job candidates from more diverse sources than traditional recruitment methods have typically tapped. One of the challenges to creating solutions to the job matching process is the outdated feel of traditional occupational interest measures, which were built under the constraints of limited technologies for implementation.

Over the past 2 years, Amavitae collaborated with a leading data scientist and an award-winning animator to reimagine traditional approaches to measuring occupational interest, with the goal of creating an engaging instrument that would appeal to a diverse set of job seekers while maintaining a high level of measurement rigor and validity. The resulting Amavitae Interest Inventory is based on Holland's (1997) six core dimensions of occupational interest. These six dimensions – Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C) – are commonly referred to by the acronym RIASEC. Prior research has identified patterns among individual's highest scoring dimensions and job compatibility. The goal of measurement instruments like the one developed by Amavitae is to link individuals to career paths aligned to their interests and personality, based on their relative RIASEC score levels.

Rather than start from scratch to develop a RIASEC measure, Amavitae selected a widely adopted instrument and reimaged it with animated visuals and a mobile device interface. The measure they chose was the O*NET Interest Profiler Short Form, which has been validated in

¹ Cheng, B. Kan, M., Levanon, G., & Ray, R. (2014). Job satisfaction: 2014 edition. New York, NY: The Conference Board.

extensive testing and is currently in wide use nationally by the US Department of Labor. Both measures -- the O*NET Interest Profiler Short Form and the Amavita Interest Inventory -- produce six scores for each individual, one per dimension. RIASEC scores are used to match individuals to appropriate careers by creating a classifier corresponding to their top three scores, in order of strength. For example, if a person's RIASEC scores were 5-3-2-7-4-2, they would be classified "S-R-E"; someone with RIASEC scores of 8-4-2-7-1-3 would be labeled "RSI". Jobs are coded to correspond to patterns of three dimensions. Once a person's RIASEC classifier is known, a system like Amavita's can suggest occupations most likely to be of interest and aligned to the individual's personality characteristics.

Purpose of the Present Study

The goal of the present study was to validate scores on Amavita's reimagined version of the O*NET Interest Profiler Short Form to ensure that individual's results on the new version would be equally effective in matching them to jobs. In a concurrent validity study, scores on the new measure are correlated against scores on a previously validated measure of the same construct -- in this case, the RIASEC dimensions. High correlations can be interpreted as meaning that the scores that individuals obtain on the new version are very close to the scores they would obtain on the traditional measure.

Given that a core goal for Amavita is to create a more engaging version of the career interest instrument, an additional purpose of the study was to determine whether or not respondents preferred the new version over the traditional one.

The Study

The validity study was carried out in two phases. The first phase was a pilot with half of the inventory items to confirm that the animated approach to item presentation held promise of

validity and reliability. Following revisions to the piloted items and development of the remaining thirty items, a final validity study phase was carried out using the entire set of items on both the Amavita^e and O*NET inventories. Because the methods were nearly identical for each study, an overview of methods is presented first, followed by a detailed description of participants and results for each phase of the study.

Overview of Methods

Participants

Participants were recruited via Amazon Mechanical Turk (MTurk), an Internet marketplace matching businesses seeking assistance with tasks to individuals wanting to complete tasks for compensation². Requesters post tasks to MTurk, specifying eligibility criteria and compensation for completed tasks. The MTurk task for the Amavita^e validity studies required workers to be classified as “master” workers within the MTurk system, and to have previously completed at least 1,000 tasks with an acceptance rate of 95% or higher. In addition, to be eligible, workers had to be between the ages of 18 and 30, and based in the United States. To confirm the former requirement, respondents were prompted to enter their age on the survey before continuing, and those who entered ages outside the eligibility range were disqualified. IP addresses were used to confirm that workers were based in the United States. Participants who completed the full task were compensated in the amount of \$3.50.

Measures

Participants completed the surveys online in a single sitting. The Amavita^e version was presented one item at a time; the O*NET was presented consistent with typical administration, with twelve items per screen and Likert-scale bubble choices. Examples of item presentation for

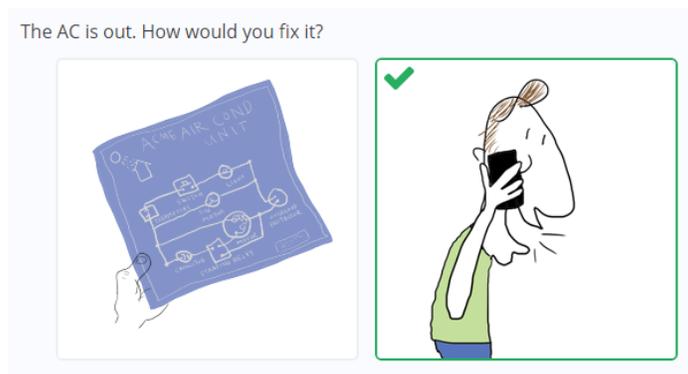
² <https://www.mturk.com/mturk/help?helpPage=overview>

both measures appear in Figure 1. Upon entering the survey, each respondent was randomly assigned to receive either the Amavitae or the O*NET version first. All respondents completed both measures. In addition to the inventories, respondents were asked basic demographic questions regarding gender and ethnicity. For the main study, additional demographic questions regarding educational level and job status were added to provide a richer description of the sample. To assess engagement levels, respondents were presented a set of questions following completion of each measure regarding its appeal, including whether they found it interesting, boring, and entertaining. After completing both measures, respondents also indicated which one they preferred.

Figure 1: *Sample Item Presentations for the O*NET and Amavitae Pilot Instruments*

How would you feel about doing each kind of work?

	Strongly dislike	Dislike	Unsure	Like	Strongly Like
Repair household appliances	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Raise fish in a fish hatchery	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conduct chemical experiments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Study the movement of planets	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Compose or arrange music	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Draw pictures	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>



One of the risks in the MTurk environment is that participants could click through without reading or attending to the prompts. To check for attention, ten of the O*NET items were

randomly selected for each participant and re-presented in random order at the end of the survey. Respondents whose responses on the repeated items were not consistent were dropped from further analysis.

Pilot Study Results

In the pilot study, the initial response sample included 225 cases. Of these, a total of 22 cases were disqualified for failing to meet the age or location eligibility requirements, for completing the task more than once, or for failing the attention check test. In the final sample of 203 participants, 56% were male and the majority (72%) reported White/Caucasian ethnicity; of the remaining 28%, 10% reported Asian/Pacific Islander ethnicity, 7% Black/African American, and 7% Latino/a/Hispanic. Ages ranged from 18-30 years old, with a mean of 27.0 and standard deviation of 2.45. Respondents were distributed over 37 states and represented all twelve of the national regions defined by the Federal Reserve.

Concurrent Validity

As shown in Figure 1, participants in the pilot study responded to Amavita items using a dichotomous scale by selecting one of two visual options; O*NET items were answered on a 5-point scale ranging from “Strongly Dislike” to “Strongly Like”, with a mid-point labeled “Unsure”. Table 1 presents the concurrent validity results for the pilot study. Scale correlations ranged from .51 to .75, with all correlations significant at $p < .01$. Reliability coefficients for the O*NET Short Form scales ranged from .85 to .92. In a 2010 study by Rounds and colleagues³, Cronbach's alpha on the O*NET Short Form ranged from .78 to .87.

³ Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2010). O*NET Interest Profiler Short Form Psychometric Characteristics: Summary and Supporting Evidence. U.S. Department of Labor National O*NET Resource Center.

Table 1: *Phase I Validity Results*

	Correlation*	O*NET α
R	0.51	0.87
I	0.75	0.92
A	0.73	0.92
S	0.72	0.90
E	0.66	0.85
C	0.56	0.94

*All correlations were significant at $p < .01$

In addition to calculating scale-level correlations, individuals' maximum scale scores on the two measures were compared for consistency. To test this consistency, the proportion of cases in which the dimension with the highest score was the same on both measures was calculated. To account for the difference in scales across the two measures, scores on the O*NET were transformed to dichotomous outcomes. The majority (78%) of participants' highest scores were on the same RIASEC dimension on both measures.

Item Analysis

To inform overall measure revisions, and to identify specific items responsible for lowering concurrent validity and maximum score agreement, an item analysis was carried out on all 30 pilot items. Figure 2 presents an example of two items from the "A" scale, one with acceptable agreement and one showing need for improvement. Agreement was calculated as the number of times a "No" or a "Yes" on the Amavitaie version was matched with a response in the "Dislike" or "Like" zone, respectively, on the O*NET response scale. "Unsure" responses were not included in the item analysis. Eight of the 30 items demonstrated agreement below 70% and were recommended for revision.

Figure 2: *Sample Item Analysis*

		ONET					
		1	2	3	4	5	Total
Amavitaes	0	22	33	15	34	18	122
	1	9	10	12	18	19	68
Total		31	43	27	52	37	190

Agreement
56%

		ONET					
		1	2	3	4	5	Total
Amavitaes	0	25	34	15	17	1	92
	1	0	0	4	59	35	98
Total		25	34	19	76	36	190

Agreement
89%

Engagement Ratings

To gauge the extent to which respondents found the Amavitaes version more engaging relative to the O*NET format, a set of engagement questions was presented following completion of each measure. As the results presented in Table 2 show, respondents rated Amavitaes as being statistically significantly more fun, more enjoyable, more entertaining, and less boring than the O*NET text version. In addition, respondents reported statistically significantly less difficulty staying focused on the Amavitaes version compared to the text version. When asked if their friends would be up for answering these questions, participants rated the Amavitaes version significantly more positively than the text version. The vast majority (92%) of respondents reported liking the format of the Amavitaes questions more than the O*NET items; and 94% said they would choose the Amavitaes version over the text version of the inventory.

Table 2: *Engagement Ratings Results*

	Amavitaē		O*NET		<i>t</i>	<i>df</i>	<i>p</i>
	Mean	SD	Mean	SD			
The questions were fun to answer.	4.4	0.8	3.3	1.2	13.69	189	<.001
The questions were easy to understand.	4.7	0.5	4.7	0.5	0.23	189	<i>ns</i>
I didn't enjoy answering these questions [R]	4.5	0.9	3.6	1.2	10.025	189	<.001
Most of these questions were boring [R]	4.5	0.8	3.3	1.3	12.471	189	<.001
My friends would be up for answering these questions.	3.8	0.9	3.2	1.0	9.704	189	<.001
I found some of the questions confusing [R]	4.5	0.8	4.7	0.5	-3.692	189	<.001
The questions were kind of entertaining.	4.4	0.8	3.1	1.3	14.836	189	<.001
It was hard to stay focused while answering these questions. [R]	4.8	0.5	4.4	0.9	5.065	154	<.001

Discussion

The pilot study results suggested that the visual approach adopted in the Amavitaē version had the potential to attain high consistency with the original O*NET version, and to do so in a more engaging and appealing format. Item analysis identified a set of specific items to revise in the final measure. Though some of the scale correlations fell within acceptable ranges (>.70), they were below the target range of .90 for strong agreement. One concern regarding attaining high agreement between the two measures was the absence of an “unsure” option in the Amavitaē items, which limited the variance of scores. An additional concern was ensuring that respondents were interpreting the items as intended in the original measure, and the need to balance a simple visual interface with the precision of meaning offered by language, e.g., written prompts.

Based on the pilot results, several changes were made to the final version of the Amavitaē measure. Items flagged for consistency issues were revised to increase comprehension. As

illustrated in Figure 3, language prompts were added to all items, and an “unsure” option was added. The response format was changed from clicking one of two images, to selecting a “Yes”, “No” or “Unsure” option, using a variety of colloquial labels (e.g., “Yeah”, “Not really”, “No clue”). The thirty additional items were developed, resulting in a complete set of items paralleling the O*NET Short Form.

Figure 3: *Example Item from Amavita Final Measure*



Main Study Results

Responses to the main study numbered 199. Of these, one respondent was disqualified for having zero variance in their O*NET responses. Those whose "attention check" responses differed by two or more points on two or more items were dropped from the analysis (n=18), as were those who had 1 or more items that differed by 3 or more points (n=4). Two respondents were disqualified because they did not complete the survey from within the United States.

The final sample of 175 participants was 61% male and 74% White/Caucasian. Similar to the pilot study, ages ranged from 18-30 years old, with a mean of 27.2 and standard deviation of

2.38. Respondents were distributed over 41 states and represented all twelve of the national regions defined by the Federal Reserve. Two-thirds (66%) held an associate’s degree or higher, and 13% reported being currently enrolled in a degree program. About two-thirds (63%) indicated that they worked a full-time job in addition to their MTurk work; 16% reported their employment consisted only of MTurk work. To inform generalizability of results, Table 3 compares ethnicity and educational attainment of main study participants to 2015 US Census data.

Table 3: *Ethnicity and Educational Attainment of Main Study Participants and 2015 US Census*

		Study Participants	US Census 2015*
Ethnicity	White/Caucasian	74%	62%
	Black/African American	9%	13%
	Hispanic/Latino/a	5%	18%
	Asian/Pacific Islander	12%	6%
Educational Attainment	Some high school	1%	12%
	Completed high school	13%	31%
	Some college or associate’s	34%	26%
	Bachelor’s	44%	19%
	Graduate	7%	11%

*Source: US Census Bureau 2015. Available at <https://www.census.gov/quickfacts/table/RHI125215/00>

Concurrent Validity

Table 4 presents the concurrent validity results for the main study. Scale correlations between the Amavitaie and O*NET versions ranged from .89 to .94, with all correlations significant at $p < .001$. Reliability coefficients for all six Amavitaie scales demonstrated acceptable values, ranging from 0.80 to 0.93; reliability for O*NET Short Form scales ranged from 0.86 to 0.94.

As in the pilot study, in addition to calculating scale-level correlations, individuals’ maximum scale scores on each measure were compared for consistency. The majority (87%) of

participants' highest scores were on the same RIASEC dimension on both versions, up 11% from the pilot study results (78%).

Table 4: *Main Study Validity Results*

	Correlation*	Amavitaα	O*NET α
R	0.91	0.80	0.88
I	0.93	0.91	0.93
A	0.94	0.89	0.93
S	0.90	0.85	0.88
E	0.89	0.81	0.86
C	0.93	0.93	0.94

*All correlations were significant at $p < .001$

Engagement Ratings

As in the pilot study, data were gathered in the main study to gauge the extent to which respondents found the Amavita α version more engaging relative to the O*NET format. Results are shown in Table 5. Respondents rated Amavita α as being statistically significantly more fun, more enjoyable, more entertaining, and less boring than the text version. Three out of four respondents said they would choose the Amavita α version over the text version of the inventory, and the majority (70%) reported liking the Amavita α item format more than the text format.

Responses on the engagement items were tested for differences by gender, education level, and employment status. Only two gender differences were found on the feedback items. Females found the O*NET Short Form significantly less enjoyable than did their male counterparts. On the Amavita α inventory, males reported statistically significantly less boredom than did females. The results for overall feedback score as well as for format preferences did not differ significantly by employment status or highest education level for either assessment.

Table 5: *Main Study Engaging Rating Results*

	Amavitae		O*NET		<i>t</i>	<i>df</i>	<i>p</i>
	Mean	SD	Mean	SD			
The questions were fun to answer.	3.95	1.0	3.40	1.0	7.391	174	<.001
The questions were easy to understand.	4.75	0.5	4.77	0.5	-0.301	174	<i>ns</i>
I didn't enjoy answering these questions [R]	4.10	1.1	3.80	1.1	3.747	174	<.001
Most of these questions were boring [R]	4.10	1.1	3.60	1.1	7.180	174	<.001
My friends would be up for answering these questions.	3.6	1.0	3.30	1.0	5.262	174	<.001
I found the questions confusing [R]	4.6	0.6	4.83	0.5	-1.620	174	<i>ns</i>
The questions were kind of entertaining.	4.0	1.1	3.09	1.1	10.481	174	<.001
It was hard to stay focused while answering these questions. [R]	4.45	0.9	4.36	0.9	1.327	174	<i>ns</i>

Order Effects

Participants were randomly assigned to the order in which they completed the Amavitae and the O*NET versions. Because presentation order could potentially have an effect on participants' engagement ratings due to fatigue or other factors, responses on all engagement items were compared between the two order groups. Nine engagement rating items were posed for each version of the RIASEC assessment, for a total of 18 items. No differences were found on the majority (15) of the 18 engagement rating scales. Respondents who received the Amavitae version first tended to rate it as more fun relative to those who completed it after they had responded to the O*NET version—mean of 4.2(sd=0.78) vs 3.8(sd=1.23), $t(172.3)=2.749$, $p=.007$. Respondents rated the Amavitae version as slightly easier—mean of 4.8(0.41) vs 4.7(0.61), $t(178)=2.304$, $p=0.22$ —if they completed it before filling out the O*NET version. Only one difference was found on engagement ratings for the O*NET version. Those who

completed the O*NET version first found it slightly more confusing than those who completed it after filling out the Amavitae version—mean of 1.2(0.59) vs 1.1(0.29), $t(137)=2.245$, $p=0.026$.

To further explore these order effects, within-group analyses were carried out (see Table 6). Both groups rated the Amavitae version significantly more fun than the O*NET version. For ease of understanding, an interaction effect was found. The Amavitae version was rated as easier to understand than the O*NET if it was completed first, but not if it was completed second. For the item regarding whether the questions were confusing, no differences were found for those who completed the Amavitae version first; those who completed the O*NET first found it relatively less confusing than the Amavitae version.

Table 6: *Engagement Rating x Order Effects Results*

	Amavitae then O*NET				O*NET then Amavitae			
	Amavitae	O*NET	<i>t</i>	<i>p</i>	Amavitae	O*NET	<i>t</i>	<i>p</i>
The questions were fun to answer.	4.2(.78)	3.4(1.03)	7.698	<.001	3.8(1.23)	3.5(1.09)	2.965	.004
The questions were easy to understand.	4.8(.41)	4.7(.50)	2.462	0.016	4.7(.60)	4.82(.41)	-2.669	.009
I found the questions confusing [R].	1.24(.61)	1.24(.60)	.000	<i>ns</i>	1.24(.66)	1.09(.28)	2.230	0.028

Order effects were also found on the two items presented at the end of the survey asking participants to compare their preference and liking of the two versions. Those who completed the Amavitae version first were more likely to prefer it to the O*NET format—mean of 4.2(1.19) vs 3.7(1.41), $t(194.2)=2.935$, $p=.004$. There was also a small difference in whether respondents would choose the Amavitae version over the O*NET if given the option—mean of 0.8(0.37) vs 0.7(0.48), $t(190)=2.943$, $p=.004$.

Importantly, analysis of data within each order group evidenced a preference for the Amavitae version over the O*NET version. Of the 97 participants who completed the Amavitae version first, 75 reported preferring the pictures to the words, and 81 of 97 (84%) said they would choose Amavitae over the O*NET if given the option. Among the 102 respondents who completed the O*NET before filling out the Amavitae version, the majority (63%) reported preferring the picture format to the text format; and two-thirds (66%) indicated they would choose the Amavitae version over the O*NET.

Limitations

Participants in the present study were recruited through Amazon Mechanical Turk. While demographic data gathered from participants suggest reasonable diversity with respect to gender, ethnicity, job status, and education levels, the results presented here are generalizable to a population with the same characteristics of the sample, and thus may have limited applicability to those not well represented within the study sample. For example, as shown in Table 3, the distributions by ethnicity and in particular by educational attainment suggest that those reporting white/Caucasian ethnicity were overrepresented relative to the national population, and that overall participants were more highly educated than is true of the nation on the whole. Though prior research is generally mixed with respect to the effects of monetary compensation on response rates and quality of responses⁴, another consideration is that participants were incentivized with compensation for their participation in the study. Notwithstanding the modest amount (i.e., \$3.50), they may have felt obliged to provide inflated responses to the engagement items.

⁴ Medway, R. L., & Tourangeau, R. (2015). Response quality in telephone surveys: Do prepaid cash incentives make a difference? *Public Opinion Quarterly*, 79(2), 524-543.

Conclusions

Results from the present concurrent validity study suggest that the Amavitaë version of the O*NET Short Form Interest Profiler produces scores that are both reliable and valid, as evidenced by strong reliability coefficients and high correlations with the original, previously validated measure. Overall, individuals scored highest on the same RIASEC dimension on both measures, an important consideration given the way that scores are used to recommend career sectors. Participants rated the Amavitaë version as more engaging relative to the traditional presentation format. This latter finding is important, since a new version of an existing measure that is merely reliable and valid would not represent a contribution to the field. The added value of the Amavitaë version is its perceived appeal and engagement qualities.

Future research should seek to replicate the present results in a more diverse sample, ideally drawn from the populations to be targeted for adoption of the Amavitaë measure. The strategy of adding language to the items yielded higher concurrent validity outcomes, though possibly at the expense of perceived appeal, given the lower preference results relative to the pilot study. Modern day technologies enable innovations in approaches to assessment, such as moving away from bubble-sheet formats towards richer visual and interactive response options. Optimizing the balance between visual and textual prompts and information, to elicit accurate responses from diverse audiences, is an area for further exploration. Further, the use of qualitative methods to probe respondents' interpretations of visual items may provide insights and ideas for improving item comprehension. Finally, the availability of rich longitudinal data from employer and recruitment databases could be leveraged over time to further validate the relation between RIASEC scores and career satisfaction and success.

Suggested Citation

Riconscente, M. M. (2016). Validation of the Amavitaë version of the O*NET Interest Inventory. San Francisco, CA: Designs for Learning, Inc.

Appendix A: Disclosure of Terms

Dr. Riconscente served as a third-party advisor and researcher on this project. She provided limited input to the Amavitaë team regarding methods for maintaining high levels of measurement rigor in the Amavitaë instrument. All data collection and analysis were carried out exclusively by Dr. Riconscente, whose compensation for conducting the research was not contingent on the outcome of the work. Dr. Riconscente has neither present nor past financial interest in Amavitaë.