



North Bay Science Discovery Day 2026 Exhibitor Evaluation “Mystery Shopper” Report

Steven M. Worker, PhD

The [2026 North Bay Science Discovery Day \(NBSDD\)](#) was a free, one-day STEM festival held at the Sonoma County Fairgrounds on March 7, 2026. The event brought together more than 100 organizations and an estimated 10,000 or more attendees for hands-on science, technology, engineering, and mathematics (STEM) learning experiences designed to spark curiosity and wonder among children.



Exhibitor Evaluation Pilot

The mystery shopper evaluation was designed as a formative quality-improvement tool, piloted at the 2026 NBSDD. Evaluators visited exhibits as regular attendees, participated briefly, then scored each activity using an 8-item rubric.

Key Takeaways

- **The mystery shopper process was a useful pilot.** The pilot generated 88 usable evaluations across 61 of 106 exhibitors (58%), providing a strong first look at exhibit quality while also showing how to refine the evaluation process.
- **Rated exhibits were generally strong.** The mean total score was 13.0 on a 16-point scale, with a median of 14.0, suggesting most evaluated exhibits met many of the rubric criteria.
- **Exhibitor strengths were clear.** Evaluators rated exhibits highest on accessibility and inclusion, friendly staff, and sparking wonder and curiosity.
- **The main improvement area was inquiry facilitation.** Lower scores appeared for asking youth to think, giving visitors agency, and making the inquiry question clear.

Methods

Mystery shopper evaluators were recruited as volunteers and completed pre-event orientation meetings. They were also provided a written orientation guide. Evaluators were instructed to participate as ordinary attendees, engage briefly with each activity, step away from the exhibit, and complete the rubric immediately. The evaluator pool included six exhibit evaluators: three adults and three high school students.

Rubric

Exhibit evaluators used an 8-item rubric to assess the quality of visitor experiences. Each criterion was scored as 2 = fully met, 1 = partly met, 0 = not met, or Not Observed. The eight criteria were summed into a total exhibit-quality scale, with a maximum possible score of 16 points. Not Observed responses were treated as missing rather than as zero, with the total score prorated to a 16-point based on the other criteria.

Caution: Inter-rater reliability was generally low to modest. This is not surprising because ratings used a short 0, 1 or 2 rubric, evaluators did not all rate the same exhibitors, and exhibit interactions can vary depending on timing, crowding, and staff member encountered. The scores are useful for formative pattern-finding, but should not be treated as precise rankings.

Rubric criterion	What evaluators looked for
1. Real materials or tools	Visitors handled real objects, materials, tools, specimens, models, or equipment rather than only viewing materials.
2. I did it myself	Visitors had agency in the activity by testing, building, measuring, controlling, recording, or doing something.
3. Clear question to explore	The activity had a visible or clearly communicated, kid-friendly question that visitors could explore in under 10 minutes.
4. Friendliness of staff	Exhibit staff welcomed visitors, invited participation, and were patient and approachable.
5. Staff get youth to think	Staff asked visitors to predict, notice, explain, compare, or reason, rather than only giving instructions or facts.
6. Scientist or engineer present	A STEM professional was clearly identified and engaged in conversation with visitors.
7. Sparks wonder and curiosity	Visitors showed signs of interest, such as asking questions, smiling, trying again, calling others over, or staying engaged.
8. Accessible and inclusive	The activity was clear for beginners, usable by different ages and abilities, physically accessible, and welcoming.

Who was rated?

The mystery shopper evaluation produced 88 usable exhibitor evaluations. Of the 106 exhibitors on the official roster, 61 exhibitors were rated (57.5%) and 45 exhibitors were not rated (42.5%). Most rated exhibitors received one evaluation. Among the 61 rated exhibitors, 41 received one evaluation, 16 received two evaluations, and three received three evaluations: Mobile Climate Science Labs, The Animal People / Lyon Ranch Therapy Animals, and The Marine Mammal Center. Medtronic received six evaluations.

By Building

Grace Pavilion received the largest share of evaluations, followed by Hall of Flowers and Garrett. E. C. Kraft was under-sampled, with only 3 of 12 exhibitors rated.

Building	Exhibitors	Exhibitors Evaluated	Total Evaluations Completed	Percent rated within building	Share of all evaluations
Grace Pavilion	41	26	38	63.4%	43.2%
Hall of Flowers	30	19	23	63.3%	26.1%
Garrett	16	8	17	50.0%	19.3%
Outdoors	7	5	7	71.4%	8.0%
E. C. Kraft	12	3	3	25.0%	3.4%
Total	106	61	88	57.5%	100%

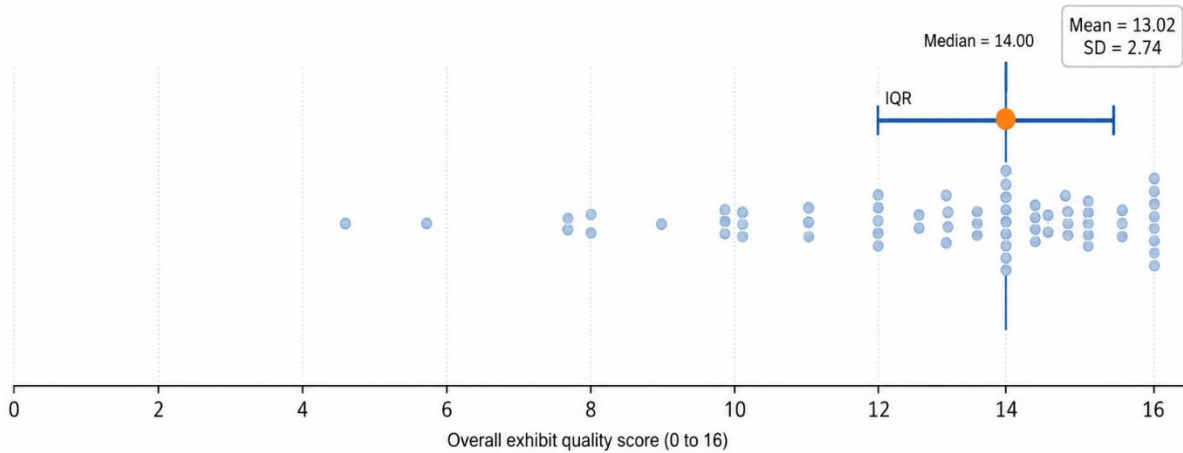


Exhibitors rated highly overall

Across rated exhibitors, scale scores were generally strong. The mean total score was 13.0 on a 16-point scale (SD = 2.7), with a median of 14.0. Scores ranged from 4.6 to 16.0, showing that most evaluated exhibits met many rubric criteria, while a smaller number may benefit from targeted coaching.

Most rated exhibits scored highly on the mystery shopper rubric

Total scale scores ranged from 4.57 to 16.00, with a median of 14.00 out of 16.



Note: Scores were averaged across raters when exhibitors received more than one evaluation.

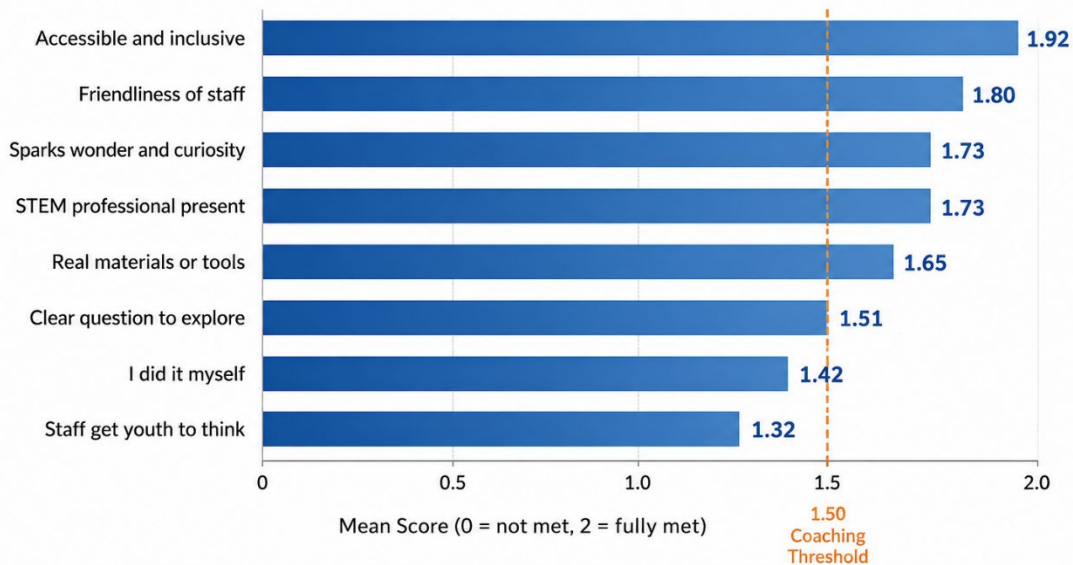


Criterion scores

Mean criterion scores were highest for accessibility and inclusion, staff friendliness, wonder and curiosity, and STEM professional presence. Scores were lower for staff getting youth to think, visitor agency, and clear inquiry question. This suggests that many exhibits were welcoming and engaging, while the strongest improvement opportunity is helping activities become more visitor-driven and inquiry-framed.

Exhibits were strongest on accessibility and staff warmth; inquiry facilitation was the clearest improvement area

Mean criterion scores on a 0 to 2 scale, where 2 = fully met.



The main opportunity is moving from hands-on activities to visitor-driven inquiry

Mean criterion scores on a 0 to 2 scale, where 2 = fully met.



WHAT THIS MEANS: Exhibits excel at being welcoming and friendly and at sparking curiosity. The biggest opportunity is helping visitors lead the experience, explore a clear inquiry question, and be challenged to think more deeply.

Best Activity Nominees

Sixteen exhibitors received at least one best-activity nomination. Evaluators highlighted hands-on interaction, visible cause-and-effect, memorable materials, robotics or technology, animal encounters, and moments when youth could see something surprising happen. These nominees provide useful examples for future exhibitor coaching because they show how short exhibit interactions can spark curiosity through visitor agency, tactile participation, and clear moments of discovery.

Exhibitor*	Best nominations (total evals)	Mean score	Evaluator feedback
NorCal FTC Champions (Mecha Mantises)	2 (2)	14.0	“Kids can see a real robot move around.”
Mobile Climate Science Labs	2 (3)	14.3	“I saw CO ₂ ppm spike in the sensor after blowing into the bottle.”
Buck Institute for Research on Aging	1 (1)	16.0	“The best moment was the result, where you could see the DNA in the tubes from the strawberries.”
Sonoma State University Chemistry Club	1 (1)	16.0	“Being able to hold a bubble. So much fun!”
The Whiting-Turner Contracting Company	1 (1)	16.0	“Amazing display. It got many kids thinking about engineering and design with their catapults.”
Sonoma County Hazmat Team	1 (1)	16.0	“They had hands-on emergency response equipment, household hazardous chemical safety, chemical reactions, acids and bases, oil clean-up, and a Level A hazmat suit for visitors to try on.”
Kiwanis Club of Santa Rosa	1 (1)	14.0	“Sparks flew when kids realized the postcard they were creating would fly on a rocket to space, then be mailed back to them postmarked ‘sent to space.’”
501st Legion (Golden Gate Garrison)	1 (1)	13.0	“The best moment of this exhibit was completing the lightsaber, and when you squeeze it, it turns on.”
NorCal Mars Society	1 (1)	12.0	“The kids seemed to love driving the rover around.”
Sonoma State University Nursing Student Association	1 (2)	15.5	“Seeing germs on my hand.”
Technology High School Medical Club	1 (2)	14.5	“Competing against a younger kid in a race to see who was better adapted to the environment.”
Pacific Gas and Electric Company (PG&E)	1 (2)	13.5	“Cool tech demos.”
Sonoma County Library	1 (2)	12.7	“Most of the children were impressed with the microscopes.”
Medtronic	1 (6)	15.1	“It made me wonder about how magic metal works in medical situations.”
The Animal People / Lyon Ranch Therapy Animals	1 (3)	13.9	“They had a number of animals, touching was encouraged, and staff were responsive to questions.”
Tesla	1 (2)	8.8	“Kids loved them. Parents loved them.”

* The table is sorted first by number of best nominations, then by strength of confirmation across evaluators, then by mean scale score.

Coaching Follow-Up Nominations

Mystery shoppers also marked whether an exhibit appeared to need coaching follow-up. These nominations should be interpreted as internal formative feedback, not as a public judgment of exhibitor quality. Overall, coaching follow-up nominations pointed less to major exhibit failures and more to common opportunities for strengthening facilitation. The most frequent themes were clearer inquiry questions, more youth control over the activity, more thinking prompts from staff, and more active invitations to participate. Some exhibits had strong materials or interesting content but relied too heavily on explanation, observation, or staff-led demonstration.

→ For 2027, the core committee could address these patterns through short exhibitor orientation materials, examples of strong inquiry questions, sample youth prompts, and targeted coaching for exhibitors whose activities would benefit from becoming more visitor-driven and inquiry-based.

Coaching theme	What evaluators noticed
More thinking questions or follow-up needed	Staff often gave instructions, explanations, or facts, but did not consistently ask youth to predict, notice, explain, compare, or reason aloud.
Increase visitor hands-on agency	Visitors sometimes handled materials, but the activity was still largely staff-led. Youth did not always test, build, measure, choose, control, or change something themselves.
Clearer inquiry question or activity purpose needed	Some activities were engaging, but the central question or purpose was not obvious to visitors within the short interaction.
More active invitation by staff	Some staff waited for visitors to approach or ask questions rather than actively greeting families, inviting youth in, and helping them start.
Clarify or identify STEM professional role	Evaluators sometimes could not tell whether a scientist, engineer, health professional, or other STEM professional was present or available to talk with visitors.
Better fit for younger children	Some activities seemed better suited for older youth or adults, while many actual attendees were younger elementary-age children.
Reduce jargon or explanation-heavy delivery	Some exhibits relied on technical language or lengthy explanation before visitors could participate.
Improve access, flow, or crowd management	Some exhibits had crowding, unclear starting points, bottlenecks, or participation barriers during busy periods.

Actions for 2027

The mystery shopper data suggests that NBSDD exhibitors are generally engaging, welcoming, and positive. I suggest the committee explore four actions for 2027.

Action 1: Help more exhibitors move to visitor-driven inquiry.

Help more exhibitors move from hands-on activities where attendees just watch, listen, or follow instructions to visitor-driven inquiry. **Visitor-driven inquiry** refers to an activity in which attendees are actively exploring a clear question by manipulating materials, making choices, testing ideas, or changing conditions, while staff support them with prompts that invite prediction, observation, explanation, and reflection. In this model, the attendee is not simply watching a demonstration or receiving information; they are driving the learning experience through hands-on investigation.

Action 2: Create a simple exhibitor quality framework

Use the criteria from the mystery shopper rubric as the core exhibit-quality framework. This would help align exhibitor recruitment, the application, orientation, coaching, and evaluation around the same expectations. The framework should be added to the Exhibitor Handbook and used in orientation materials.

Action 3: Build a coaching pipeline

The coaching process should be supportive. Most exhibits with coaching needs appeared to have good ideas that could be strengthened with clearer facilitation, stronger youth agency, or a more visible inquiry question.

Tier	Exhibitor group	Committee response
Recognition	Best-activity nominees and high-scoring exhibitors	Feature them in communications and invite them to model for other exhibitors.
Light coaching	Returning exhibitors with good activities	Send 2 to 3 practical suggestions before 2027.
Required consultation	New exhibitors, low-scoring exhibits, or exhibits with repeated concerns	Schedule a conversation before acceptance or final confirmation. Focus on making the activity hands-on, visitor-driven, and appropriate for younger attendees.

Action 4: Improve the 2027 mystery shopper evaluation design

- Assign mystery shoppers to specific buildings or zones and create a target list so every exhibitor receives at least one rating.
- Define “clear inquiry question” more concretely, since this was one of the least consistent criteria. Inquiry questions appeared in the printed program, but they were not always visible at the exhibit itself. Exhibitors should be encouraged to display or clearly state their inquiry question at the booth.