

# REPORTING WITH AN EVALUATOR AUDIENCE IN MIND



## A BISE Synthesis Paper

**By Amy Grack Nelson & Zdanna Tranby**

Department of Evaluation and Research on Learning



© September 2015

The BISE Project is funded by NSF award #1010924

A special thanks goes out to the Science Museum of Minnesota staff that spent many hours helping to refine the code definitions and code the 520 reports included in the BISE project: Gayra Ostgaard Eliou, Gretchen Haupt, Al Onkka, Dan Bernstein, Arden Ashley Wurtmann, Scott Van Cleave, Denise Deng, and Zdanna Tranby. We'd also like to thank all of the evaluators who posted their evaluation reports on [informalscience.org](http://informalscience.org).

Please direct any questions or comments about this paper to Amy Grack Nelson, Evaluation & Research Manager, at the Science Museum of Minnesota 651-221-4575 or [agnelson@smm.org](mailto:agnelson@smm.org).

# ABSTRACT

There are a number of places evaluators can share their reports with each other, such as the American Evaluation Association's eLibrary, the website [informal.science.org](http://informal.science.org), and organizations' own websites. Even though opportunities to share reports online are increasing, the evaluation field lacks guidance on what to include in evaluation reports meant for an evaluator audience. If the evaluation field wants to learn from evaluation reports posted to online repositories, how can evaluators help to ensure the reports they share are useful to this audience? This paper explores this question through the analysis of 520 evaluation reports uploaded to [informal.science.org](http://informal.science.org). The researchers created an extensive coding framework to align with features of evaluation reports and evaluators' needs. It was used to identify how often elements were included or lacking in evaluation reports. This analysis resulted in a set of guiding questions for evaluators preparing reports to share with other evaluators.

# INTRODUCTION

In recent years, the evaluation field has turned its focus to improving reporting practices through rethinking report layouts and the visualization of data (Azaam & Evergreen, 2013a, 2013b). Improving how evaluators present data is an important part of making reports useful for various audiences, but what else needs to be considered? A report is only as useful as the information contained within its pages and striking graphics can't make up for a report that may be missing important details of a study.

But what information is key to include in evaluation reports for various evaluation stakeholders? When writing evaluation reports, evaluators tailor their reports for their different audiences. The literature provides guidance on what these reports might look like for primary intended users and other key stakeholders of an evaluation, but what about when sharing reports with other evaluators? There are currently a number of places evaluators can share their reports with each other, such as [informal.science.org](http://informal.science.org), the American Evaluation Association's eLibrary, and organizations' own websites. As the use of these sites continues to grow and others become available, how do we make sure the reports posted to these sites are useful to evaluator audiences?

Within the evaluation field, there are a number of resources that provide guidance for writing reports, although many of these are focused on reports for individuals who are clients, or primary audiences, of evaluations (American Evaluation Association, 2004; Fitzpatrick et al., 2011; Miron, 2004; Patton, 2012; Torres, 2005; Torres, Preskill, & Piontek, 2005; Yarbrough, Shulha, Hopson, & Caruthers, 2011). But what about when evaluators share their reports with evaluation professionals, not a primary audience? Journal articles provide guidelines for reporting, but evaluators typically share their studies through an evaluation report, not a journal article. Guidance is lacking on what to include in evaluation reports that are meant for an evaluator audience. However, guidelines and resources focused on evaluation reporting in general can serve as a starting place.

## Evaluation Reporting

How do evaluators determine what information is useful to their various reporting audiences? First, they must identify stakeholders and their information needs. What might be useful to one stakeholder may not be useful to another. In the *Encyclopedia of Evaluation*, Torres (2005) describes three distinct audiences for evaluations: primary, secondary, and tertiary. Primary audiences are often the individuals who fund or run the program and will use the results to make decisions about the project. This audience would also be considered the primary intended users of the evaluation (Patton, 2012). Secondary audiences have limited direct involvement with running the project but are still interested in the evaluation report because the results may affect them in some way. This audience includes program participants and staff in the organization. Tertiary audiences may or may not have a connection to the program, but they are still interested in hearing about the program and seeing the results of the evaluation. This audience includes the public and professionals (such as other evaluators).

Fitzpatrick, Sanders, and Worthen (2011) also discuss reporting interests of audiences, but from a slightly different perspective. They stress that some stakeholders may be "methodologically oriented" and wish to see detailed descriptions of the study's methods, analyses, and other technical details (p. 458), whereas other audiences may not be as interested in the technical aspects of the evaluation as they are in the findings. Knowing an audience's technical interests can help tailor the level of detail to include in reports.

Some resources offer guidance around the key elements to include in an evaluation report. Fitzpatrick et al. (2011) provide an outline of what should be included in any “formal, final evaluation report” (p. 469). Key report elements include purpose of the evaluation; description of the evaluation object, or evaluand; evaluation questions; overview of the evaluation plan; data collection methods; sampling procedures; and data analysis techniques. Although they suggest what to include in a report, they stress that all of the information does not need to be in the body of the report but could be included in an appendix for individuals who are interested in more technical details. Miron (2004) provides guidance on what could be included in various sections of an evaluation report, but acknowledges that some report elements may not apply to certain audiences. Some of the report elements in Miron’s Evaluation Report Checklist (2004) relating to the background and methodology of an evaluation include evaluation purpose and questions, description of the project being evaluated, evaluation approach or model being used, evaluation design, sample size, data collection methods, descriptions of data collection instrument(s), and data sources.

Other resources in the evaluation field stress the importance of adequate reporting to ensure there is an accurate representation of the evaluation process and findings. The American Evaluation Association’s (2004) Guiding Principles provide guidance related to communicating evaluation in the Systematic Inquiry Principle A3:

Evaluators should communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret and critique their work. They should make clear the limitations of an evaluation and its results. Evaluators should discuss in a contextually appropriate way those values, assumptions, theories, methods, results, and analyses significantly affecting the interpretation of the evaluative findings. These statements apply to all aspects of the evaluation, from its initial conceptualization to the eventual use of findings.

The Joint Committee on Standards for Educational Evaluation’s Program Evaluation Standards address communication and reporting in Accuracy Standard A8: “Evaluation communication should have adequate scope and guard against misconceptions, biases, distortions, and errors” (Yarbrough, Shulha, Hopson, & Caruthers, 2011, p. 217). The Standards also stress the need to communicate clearly enough to “prevent misunderstandings and misconceptions about information sources, data collection methods, and alternative approaches” (p. 220).

## Evaluators as an Audience

As illustrated above, the evaluation literature provides general reporting guidance, but lacks guidance specific to evaluator audiences. So what happens when evaluators become an audience of reports that weren’t written with them in mind? How well are evaluators’ needs met? These questions and, as a result, a reflection on the nature of writing reports for an evaluator audience, became an unintended outcome of the National Science Foundation (NSF) funded Building Informal Science Education (BISE).<sup>1</sup>

In 2010, the BISE project brought together a team of evaluators and researchers to investigate how a rich collection of over 500 evaluation reports voluntarily posted to the website [informal.science.org](http://informal.science.org) could be used to advance understanding of the trends and methods used in the evaluation of informal education projects. The evaluators working on the BISE project produced an extensive coding framework for analyzing the reports and commissioned five authors to produce synthesis papers that examined different aspects of informal science education evaluation. As the BISE team members

---

<sup>1</sup> The Building Informal Science Education project is a collaboration between the University of Pittsburgh Center for Learning in Out of School Environments, the Science Museum of Minnesota, and the Visitor Studies Association.

began to read the reports and develop a coding framework, they noticed some reports were lacking what they felt was information necessary to understand how an evaluation was carried out and put evaluation findings into context. As the BISE project team began coding reports, synthesizing findings, and sharing the coding framework with others, an overarching question about evaluation reporting practices emerged. If the informal learning field want to learn from evaluation reports on informalscience.org and similar types of repositories, how can evaluators help to ensure the reports they post are useful to these audiences? This article answers this question through the analysis of what was included and lacking in evaluation reports on informalscience.org. The article ends with guiding questions for evaluators to ask themselves when preparing reports to share with other evaluators on online repositories such as informalscience.org.

## METHOD

### Sample

Informalscience.org houses a diverse range of evaluation reports that individuals voluntarily post to the site. Reports range from evaluations of exhibitions to afterschool programs to media. The sample for this article was selected from all of the evaluation reports posted in the “Evaluations” section of the site to June 1, 2013. This includes reports written between 1990 and 2013. The reports are primarily related to informal science education projects, but anyone evaluating informal education projects is encouraged to share their work on the site, so the sample also includes projects covering topics such as history and art. To be included in the sample, the report had to be an actual evaluation report. There were some instances where literature reviews, research reports, and executive summaries were posted under the category of “Evaluations” on informalscience.org. These documents were excluded from the sample. A total of 520 reports fit the sample criteria.

At the time the sample was selected, there were no guidelines on informalscience.org to suggest what report elements should be included in a posted evaluation report. Additionally, many of the reports in the sample may not have been posted to informalscience.org with an evaluator audience in mind. During conversations with informal education evaluators, the BISE team found that many reports posted on informalscience.org were what evaluators provided to their clients, and the reports hadn’t been altered in any way before posting. For these reasons, the formats of the reports and what was included in them varied widely.

### Coding Process

Coding categories and related codes were created to align with key features of evaluation reports and the potential coding needs of BISE synthesis authors and BISE database users. Various resources were used to help inform the development and operationalization of these codes. As previously discussed, the evaluation reporting literature was consulted. Additional guidance came from the Center for the Advancement of Informal Science Education (CAISE) Portfolio Inquiry Group’s codebook, which was developed to analyze publicly available NSF informal science education (ISE) award data (Baek, 2011). The BISE project team also referred to the categories used in the NSF ISE Online Project Monitoring System’s Baseline and Annual Surveys, which NSF principal investigators complete to report on their project and related evaluation activities (National Science Foundation, 2011a, 2011b). During the development of the initial codes, a BISE team member conducted preliminary coding of a sample of 230 reports on informalscience.org. This process helped to refine codes and identify additional codes that emerged from the data. The codes were further refined

based on feedback from evaluators during a presentation and discussion at the 2011 Visitor Studies Association conference.

The final BISE coding framework had 32 different coding categories (Building Informal Science Education Project, 2014). The 10 categories discussed in this article are included in Table 1. A detailed list of the codes under each coding category in Table 1 is included in Table A1 in the Appendix.

TABLE 1. DEFINITIONS OF THE BISE CODING CATEGORIES INCLUDED IN THIS PAPER

Coding category	Definition
Evaluand	The object(s) being evaluated.
Project setting	The location of the project being evaluated, such as an institution, home, or type of media.
Evaluation type	The type of evaluation, such as formative or summative.
Evaluation purpose	The focus of the evaluation, which includes the evaluation questions.
Data collection methods	The method(s) used to collect data for the evaluation.
Type of instrument(s) included in the report	The type of data collection instrument(s), if included with the report.
Sample for the evaluation	The types of individuals that composed the sample or samples of the evaluation, such as the general public, specific age groups, or adult-child groups.
Sample size	The description of sample size for each data collection method.
Statistical tests	The statistical tests used for the data analysis.
Recommendations	The inclusion of evaluator-generated recommendations based on the evaluation data.

Once coding categories were established, the BISE project team at the Science Museum of Minnesota worked to establish inter-rater reliability. Two or three coders, depending on the coding categories, independently coded the same 30 reports that had been randomly selected from the sample. If percent agreement was below 90% for a coding category, the coders discussed and resolved coding disagreements, revisited the coding definitions, and revised code definitions as necessary. The coders would then proceed to code another 30 randomly selected reports. This process was repeated until at least a 90% percent agreement was reached for each coding category (based on recommendations from Miles & Huberman, 1994). The mean percentage of agreement across all variables was 95%, and ranged from 90%-100%. Once adequate agreement was reached between coders, coding and analysis took place in the qualitative analysis software NVivo 9.

## RESULTS

All 520 reports were coded for each of the coding categories and related codes listed in Table 1 and Table A1. What follows is an overview of the findings separated by coding category.

### Evaluand

All of the reports described the evaluand in such a way that it was easily categorized into one of the evaluand codes. Evaluand was also the only report element, or coding category, that coders found in all 520 reports and did not have trouble identifying.

## Project Setting

Almost all of the reports (98%) adequately described the setting in which the project took place (e.g., museum, community site, school, zoo, park, library, on the computer). There were 11 reports that provided insufficient information for the team to accurately identify and code the project setting. For example, one report evaluated teen programs without saying where the programs were held. Some reports described the state or region where a project may have taken place but not the setting where the project actually occurred.

## Evaluation Type

Nearly all of the reports (99%) described the type of evaluation carried out. Only five reports lacked sufficient information to easily categorize the evaluation type. One report was termed an exploratory study. Since the authors did not describe what they meant by “exploratory,” it was difficult to determine what kind of evaluation it was. The content being evaluated wasn’t up for revision, which suggests a summative evaluation, but the word “exploratory” has more of a formative or front-end connotation. Another report went back and forth between calling the work a research study and an evaluation without providing specifics about the type of evaluation being conducted. The other three reports failed to describe the evaluation in a way that was clear what type of evaluation was conducted.

## Evaluation Purpose

Most evaluation reports (92%) described the purpose of the evaluation. This meant that the authors went beyond stating the type of evaluation to also describe the evaluation’s focus and/or include the overarching evaluation questions for the study. However, 8% of evaluation reports lacked information about the purpose of the evaluation. Of the 479 reports that described the purpose of the evaluation, less than half (47%) included evaluation questions when doing so.

## Data Collection Methods

Data collection methods were described in almost all of the reports (98%). Only 10 reports did not adequately describe the data collection methods used for the evaluation. For example, in one instance the term “user testing” was mentioned, but the report didn’t explain what this meant in terms of the data collection methods used. Some reports said they asked individuals for feedback but didn’t indicate the method in which feedback was obtained.

## Data Collection Instruments

Less than half the reports (44%) included at least one data collection instrument. Of the 227 reports that had instruments, the most frequently included instruments were surveys (69%) and interviews (60%). The reports with instruments also included observation protocols (22%), focus group protocols (14%), timing and tracking forms (14%), and/or other types of instruments (8%) such as log books, journals, and professional critique forms.

## Sample for the Evaluation

The sample for the evaluation was coded based on whether the evaluation’s sample was the general public (a mix of adults and children), specific age groups, or adult-child groups. Over half of the evaluation reports (59%) described the sample for each of the data collection methods used in the study. However, this means that over two-fifths of the reports (41%) did not describe the sample for

at least one of the data collection methods. A common example of this is when authors describe their sample using terms such as “visitors,” “users,” “viewers,” or “community members,” but never describe how they are specifically defining those terms.

Most of the reports (87%) indicated the sample size for each type of data collection method that was carried out for the evaluation. The rest of the reports either failed to include the sample size for one or more of the methods in a study (11%) or did not include the sample size for any of the data collection methods (3%). In some instances where sample size was missing, evaluators reported how many people responded to an individual survey or interview question, but never described the overall sample size for the survey or interview.

## Statistical Tests

Findings from statistical tests were reported in half of the reports (50%). Of the 260 reports that included statistical tests, over a third (35%) stated that their results were “significant” but did not indicate what tests were run to reach this conclusion. In some cases, it was unclear if “significant” meant the findings were statistically significant.

## Recommendations

Evaluators included recommendations in a majority (59%) of the evaluation reports. Table 2 illustrates the percentage of reports of each evaluation type that include them. Overall, remedial, front-end, and formative evaluations had the highest occurrences of recommendations.

TABLE 2. PERCENT OF REPORTS THAT INCLUDE RECOMMENDATIONS BY EVALUATION TYPE

Evaluation type	Percent that included recommendations
Front-end ( <i>n</i> = 62)	69%
Formative ( <i>n</i> = 104)	63%
Remedial ( <i>n</i> = 7)	100%
Summative ( <i>n</i> = 341)	56%
Audience study ( <i>n</i> = 14)	50%

# DISCUSSION

As described in the results, the information included in the evaluation reports of informal learning experiences was highly variable. Some elements were frequently included in reports posted to [informal.science.org](http://informal.science.org), while other elements were sometimes missing. The evaluand was the only report element clearly described in all 520 reports. Almost all of the reports described the project setting, the type of evaluation, and the data collection methods used in a study. Although the evaluation purpose (92%) and sample size (87%) were included in a majority of the reports, they were both features that some reports were missing, and, upon looking more closely at the evaluation purpose, less than half (47%) included the evaluation questions related to the purpose. Report elements that were most often lacking were the data collection instruments, description of the sample for the evaluation, description of statistical tests used, and inclusion of recommendations.

Data collection instruments were the most frequently absent report elements, which were missing in over half (56%) of the evaluation reports. Data collection instruments are valuable tools other

evaluators use to understand how evaluation data are gathered in informal settings. Instruments, such as surveys and interview protocols, help readers understand what questions were asked of the evaluation sample, how the questions were asked, and thus how the reported results relate to the data collection method. Posting reports without instruments can limit the report's utility for other evaluators. For instance, some reports that did not include the instruments failed to describe the survey or interview questions asked of their sample. This makes it difficult for the evaluator reading the report to understand how the data were obtained or to judge the quality of the analyses.

Over two-fifths of reports (41%) did not describe the sample for at least one of the data collection methods. A common occurrence was the use of vague descriptors for the sample, such as "users" and "visitors," without elaborating on what those terms actually meant. This makes it difficult to interpret an evaluation's results and understand what groups the findings apply to.

Of the reports that used statistics, over a third (35%) said that results were "significant," but did not specify what test was used. Without indicating the test, the reader cannot judge the appropriateness of the statistical test, the quality of the analysis, or accuracy of the interpretations. When the reader knows what statistical analysis was conducted, he/she can better engage with the data and arguments presented in a report. Another issue was the use of the word "significant" on its own, instead of saying "statistically significant." If a report states that results were "significant" instead of saying "statistically significant," readers cannot be certain if the significance is actually related to a statistical finding.

Recommendations were not included in over two-fifths of the evaluation reports (41%), and their occurrence varied by type of evaluation. Evaluators have varying philosophies about providing recommendations to clients, and these philosophies may vary depending on the type of evaluation being carried out. There is no way to know from the evaluation reports if recommendations were actually generated as part of the evaluation process; all that is known is how often they were reported. If evaluators do provide recommendations as part of an evaluation, including them in reports they post to repositories such as [informal.science.org](http://informal.science.org) can be useful to evaluators, as well as practitioners. For instance, the usefulness of recommendations for the informal learning field is exemplified in a BISE synthesis paper that examines what lessons can be learned about exhibition development from synthesizing recommendations in exhibition summative evaluation reports (Serrell, 2014).

## Limitations

This paper sheds light on the quality of reporting practices of informal science education evaluators and highlights areas where reporting practices could be improved. There are a few limitations to keep in mind, however, when reviewing the findings.

A major limitation of this paper is that the sample of reports posted to [informal.science.org](http://informal.science.org) may not be representative of the reporting practices of all evaluators of informal education projects or the evaluation field at large. The findings are limited to the convenience sample of reports that evaluators voluntarily posted on [informal.science.org](http://informal.science.org) up until June 1, 2013. However, even with this limitation in mind, no other resource like [informal.science.org](http://informal.science.org) exists. The number and diversity of reports offers a unique window into how evaluations are conducted and reported on in the informal learning field.

A second limitation of this paper is the lack of discussion around how evaluators are reporting on a project's subject area(s). The coding category "subject area" was originally in the BISE coding framework but was removed when the BISE team was unable to reach adequate percentage

agreement after iterative cycles of revisions to the subject area definitions and recoding the data. (The highest percent agreement reached for subject area was 73%). The difficulties in coding were a result of many reports lacking sufficient information about the object being evaluated for coders to accurately distinguish the subject area or areas (e.g., physical science, engineering, health) of the project. Although this is a limitation, it is also an important finding for evaluation reporting practices in informal learning, as knowing the subject area of a project is important if evaluators are looking for how projects covering a similar topic are evaluated.

## CONCLUSION

With the increase in avenues for sharing evaluation reports online, how can evaluators help to ensure the reports they post are useful to other evaluators? By exploring what is available in reports posted on [informal.science.org](http://informal.science.org) and reflecting on reporting practices, this article brought to light where reports in the field of informal education evaluation may already meet the needs of an evaluator audience and where reports are lacking. To help maximize learning and use across the evaluation community, below are some guiding questions evaluators can ask themselves as they prepare a report meant for an evaluator audience.

### GUIDING QUESTIONS FOR PREPARING REPORTS FOR AN EVALUATOR AUDIENCE

1. Have I described the project setting in a way that others will be able to clearly understand the context of the project being evaluated?
2. Is the subject area of the project or evaluation adequately described?
3. Have I identified the type of evaluation?
4. Is the purpose of the evaluation clear?
5. If I used evaluation questions as part of my evaluation process, have I included them in the report?
6. Have I described the data collection methods?
7. If possible, can I include data collection instruments in the report?
8. Do I provide sufficient information about the sample characteristics? If I used general terms such as “visitors,” “general public,” or “users,” do I define what ages of individuals are included in that sample?
9. Have I reported sample size for each of my data collection methods?
10. If I report statistically significant findings, have I noted the statistical test(s) used? Do I only use the word “significant” if referring to statistically significant findings?
11. If I provided recommendations to the client, did I include them in the report?

These questions can serve as a checklist for any evaluator preparing a report to share with their evaluator colleagues. The simple addition of a few technical details to a report narrative or a longer technical appendix can help maximize the contribution an evaluation report to the informal education evaluation community and the larger evaluation field.

# REFERENCES

- Azaam, T., & Evergreen, S. (Eds.). (2013a). Data visualization: Part 1. *New Directions for Evaluation*, 139.
- Azaam, T., & Evergreen, S. (Eds.). (2013b). Data visualization: Part 2. *New Directions for Evaluation*, 140.
- American Evaluation Association. (2004). *American Evaluation Association guiding principles for evaluators*. Retrieved from <http://www.eval.org/publications/guidingprinciples.asp>
- Baek, J. Y. (2011, March). CAISE portfolio inquiry group: Codebook. Paper presented at the CAISE Informal Commons Working Group, Corrales, NM.
- Building Informal Science Education Project. (2014). *The Building Informal Science Education (BISE) project's coding framework*. St. Paul, MN: Science Museum of Minnesota. Retrieved from [www.visitorstudies.org/bise](http://www.visitorstudies.org/bise)
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, NJ: Pearson Education.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Miron, G. (2004). *Evaluation report checklist*. Kalamazoo, MI: Western Michigan University.
- National Science Foundation. (2011a). *Informal Science Education baseline submission*. Washington, DC: Author.
- National Science Foundation. (2011b). *Informal Science Education annual update*. Washington, DC: Author.
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- Serrell, B. (2014). *A review of recommendations in exhibition summative evaluation reports*. Chicago, IL: Serrell & Associates. Retrieved from <http://visitorstudies.org/projects/bise>
- Torres, R. (2005). Reporting. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 371-376). Thousand Oaks, CA: Sage.
- Torres, R. T., Preskill, H. S., & Piontek, M. E. (2005). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations* (2nd ed.). Thousand Oaks, CA: Sage.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

# APPENDIX

TABLE A1. CODES UNDER EACH CODING CATEGORY OF THE BISE CODING FRAMEWORK (BISE, 2014)

Coding category	Codes
Evaluand	Audience study, Broadcast media, Collaboration or partnership, Conference, Educational materials, Event or festival, Exhibition, Forums and science cafes, Mobile or handheld technology, Out-of-school time program, Planetarium show, Professional development, Public participation in scientific research program, Public programming, School-related programming, Volunteer program, Website or software, Other
Project setting	Aquarium or zoo, Arboretum or botanical garden, Art museum, Children's museum, Community site, Computer, General museum, History museum, Home, Library, Movie theater, Natural history and/or anthropology museum, Park and/or nature center, Planetarium, Radio, School, Science-technology center or museum, Television, University or college, Don't know
Evaluation type	Audience study, Front-end, Formative, Remedial, Summative, Don't know
Evaluation purpose	Included evaluation questions; Described purpose, but no questions; Did not describe purpose or questions
Data collection methods	Artifact review, Card sort, Comment cards or books, Concept map, Drawings, Focus group, Interactive methods, Interview, Journals, Observation, Participation data, Professional critique, Recorded conversation, Survey, Timing and tracking, Web analytics, Other, Didn't describe data collection methods
Type of instruments provided in the report	Focus group protocol, Interview protocol, Observation instrument, Survey instrument, Timing and tracking instrument, Other type of instrument, No instruments provided
Sample for the evaluation	Sampled general public, Sampled specific age groups (which was further coded based on the ages of the children or adults), Sampled adult-child groups, Don't know who was sampled
Sample size	Indicated sample size for all data collection methods, Indicated sample size for some data collection methods (not all), Did not indicate sample size for any data collection methods
Statistical tests	ANOVA, Chi-square, Coefficient alpha, Correlation, Effect size, Fisher's exact test, Hierarchical linear model, Kappa coefficient, Kruskal-Wallis test, Mann-Whitney U test, McNemar test, Regression, T-Test, Wilcoxon signed-rank test, Other, Said "significant" but didn't indicate test, No test
Recommendations	Recommendations included, No recommendations