

APPENDIX I: Quality Control Procedures


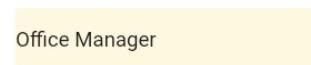








CEC-200-2023-017-API

This appendix describes quality control (QC) procedures applied to survey data.

Stage 1: Data Validation through CEUS Tool™

The CEUS Tool™ has automated QC checks that assist the surveyor during data entry. The automated data field checks highlighted data fields to identify data gaps, inconsistencies, and values that are out of range that the surveyor needs to address. Data entry spaces are initially formatted with broad red diagonal cross hatching, and change to green underlined text as a value in the acceptable range has been entered. Examples of the formatting and symbols are provided in Table 1.

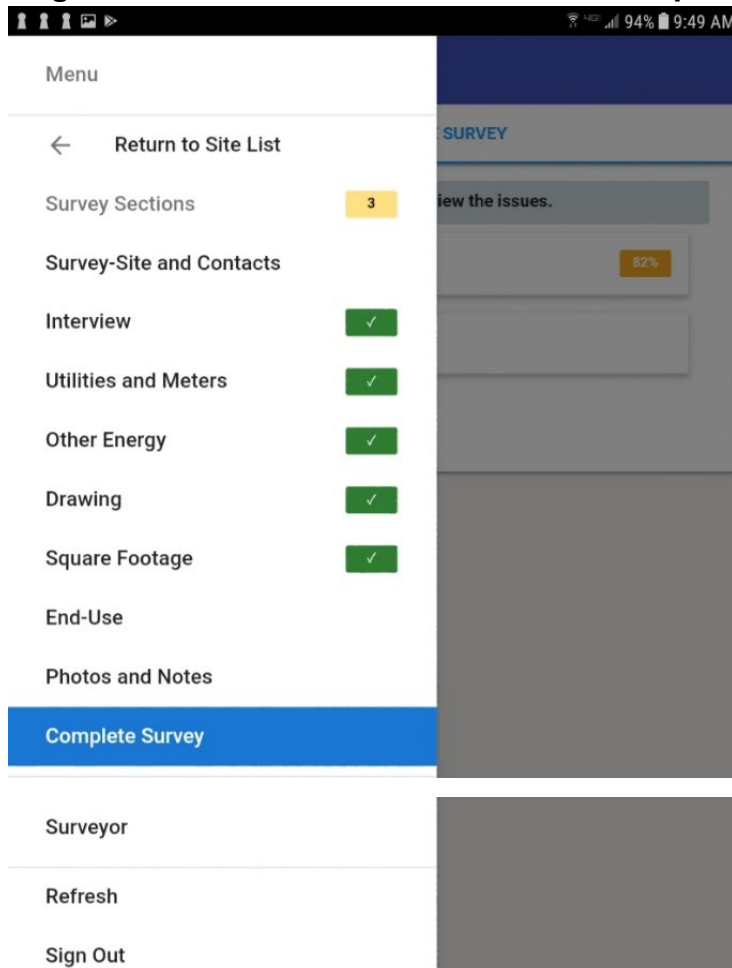
Table 1: CEUS Tool™ Entry Display Formats and Descriptions

Tablet Display Format	Description
Title 	Field that must be completed
Title Owner 	String Field that has been entered (green underline)
Title Office Manager 	Field that is actively being entered (blue underline)
Solar 	Optional entry field or displayed only (gray underline)
Gas 100 	Numeric field entered with value in accepted range
Electricity 200 	Numeric field entered with value outside of accepted range
Type 	Dropdown list not yet selected
Type HVAC 	Dropdown list after selection
 Reports to this location?	Yes/No check box not yet answered
 Reports to this location?	Yes/No check box, checked = Yes
 Reports to this location?	Yes/No check box, unchecked = No
 Can't determine year	Optional check box (not needed)
 Can't determine year	Optional check box (answered)

Field entry values outside a predefined range generate a red underline and broad red diagonal cross hatching. For example, if the sum of certain entries which was supposed to be 100% does not add-up to 100%.

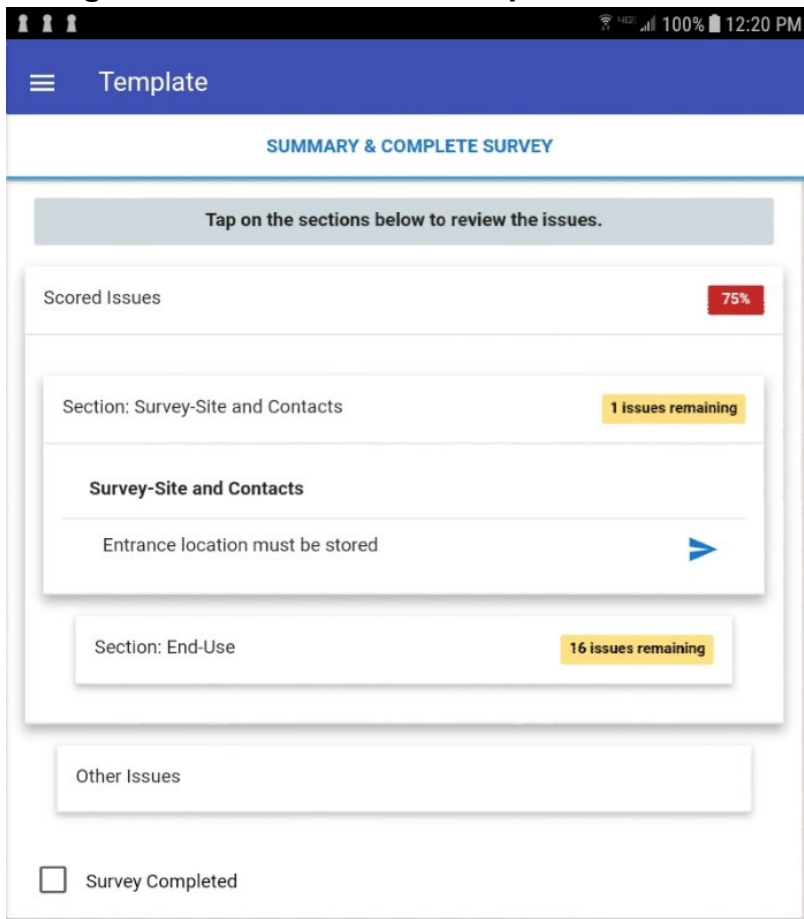
The CEUS Tool™ has progress indicators for the surveyor. As tab sections in the Tool are completed a green check mark is placed next to the tab name in the dropdown list of tabs, see example in Figure 1 showing five completed tabs. The number of sections remaining to be completed is shown in the small yellow box, in this example three more sections need to be completed. Once all the tab names have a green check mark the survey is considered as completed and can be submitted.

Figure 1: CEUS Tool™ Tab Checklist Dropdown Showing Completed Tab Sections.



The incomplete or missing questions in each of the remaining unchecked tabs (without green check mark) are listed in “Complete Survey” tab shown in Figure 2. In this example, the initial QC score is 75% and is highlighted in red to draw attention that it was less than 90%.

Figure 2: CEUS Tool™ Example List of Questions Remaining to be Completed.



During the course of the survey, the Tracking and Improvement Group identified several automated checks from Stage 2 QC that could be readily coded into Stage 1 QC. These QC checks enhanced Stage 1 QC efficacy, which ultimately reduced the number of sites that were escalated to Stage 3.

Stage 2: Automated QC Routines

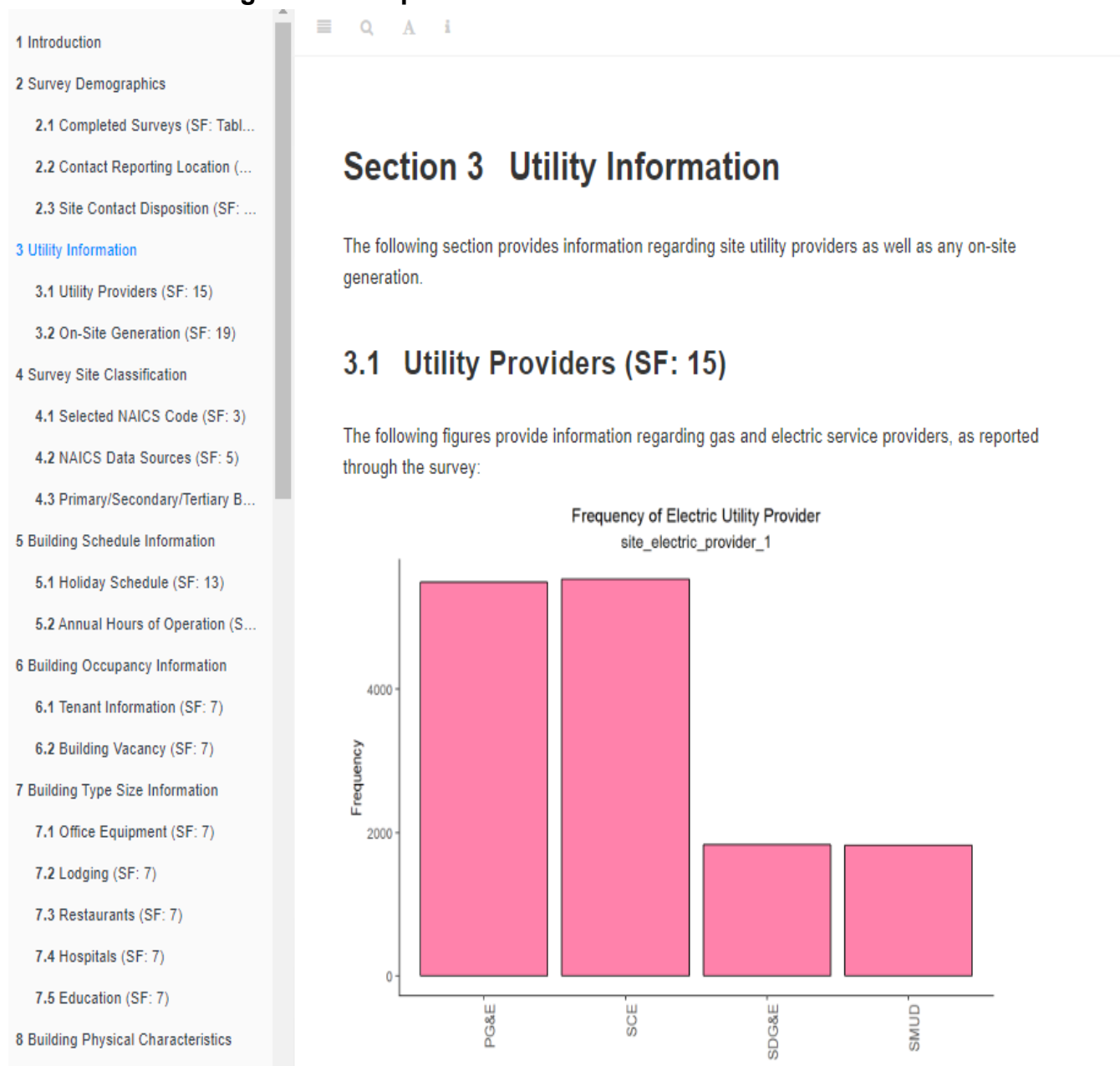
Site survey data was periodically extracted from the CEUS Tool™ and run through an automated routine scripted in the R programming language. The automated routine consisted of two different steps and applied 101 distinct QC checks, jointly developed by ADM and Energy Commission staff:

1. The first step checked values in the data set for data quality, external validity, and internal validity. Checks for data quality included logical checks to ensure that data responses were in the appropriate data format. For example, Boolean fields that were intended to have responses coded as 1 or 0 were reviewed to ensure that other values were not recorded. External validity checks included checking to make sure that reported values were theoretically possible. For example, cases of construction year may have had values reported that were not theoretically possible, such as construction years later than the year 2019. Internal validity checks included checking to make sure that responses to questions that were related to one another had responses that agreed with one another. For example, the total square-footage of a whole site was compared to the sum of the square-footages of the tenants belonging to the site in order to

ensure that the sum of the tenant square-footage did not exceed that of the whole site. A QC failure report was generated based on these logical checks, and all sites that failed QC were then submitted for manual review.

2. The second step generated an HTML dashboard that visualized the unweighted responses via frequency charts, bar charts, histograms, and summary tables. These dashboards were submitted to the Commission, as well as reviewed by ADM staff, to identify additional logical filters to add to the first step in the automated QC routine. For each numeric field in the survey, a bar chart of means by building-type was generated, as well as histograms segmented by building-type. For each Boolean field in the survey, a bar chart of means by building-type was generated. For each categorical field in the survey, frequency charts were created. Figure 3 provides a screen shot of the HTML dashboard.

Figure 3: Sample Screenshot of the HTML Dashboard



Stage 3: Manual QC Review

The third level of the QC consists of a manual review of sites that were previously identified during the Stage 2 automated QC routine. ADM QC staff reviewed issues identified in the automated QC report and determined the nature of missing or flagged items. ADM then rectified the issues using one of four possible methods:

1. ADM QC staff determined that anomalous responses were accurate, and no further action was taken on the site.
2. The items were submitted back to the original surveyor for correction or corrected by a QC team member.
3. A follow-up telephone survey was done to collect additional information.
4. Additional information could not be collected, and the site was either discarded and a new site selected, or missing fields remained in the site's data and the site was skipped during analysis of missing fields.

The steps taken to remediate sites during the QC process are described in further detail in the following section.

Remediation

In cases where sites required remediation, many issues and their solutions could be found through a simple Google search. Most of the problems with physical building-type, the number of buildings, or the site's HVAC could be immediately identified using Google satellite views. Further issues with the building, such as the existence of exterior lighting or even the existence of gas meters, could be addressed by viewing the site's street view.

Often the main business activity and NAICS code could be verified by finding the business's website, Yelp and Google reviews, or Manta description. Pictures uploaded by both customers and site owners on Yelp and Google often also provided the ability to verify and correct issues with energy using equipment, i.e. refrigeration, cooking equipment, office equipment, etc. Finally, outliers in operational hours, employee count, and occupancy could often be found through data provided by Google about sites' schedules and busy times throughout the week. Around 5% of sites had little to no internet footprint which made remote research nearly impossible. However, the vast majority of sites, particularly sites above stratum 1, provided a lot of data about themselves online.

When the necessary information for remediating a site was not available online, follow-up contact with the original site-contact was conducted via telephone call. This method was particularly effective when the missing information could be provided through documentation such as utility bills, solar installation reports, propane purchasing receipts, etc. Additionally, although site-contacts were sometimes unable to provide answers to the surveyor at the time of the appointment, follow-up rounds of contact allowed site-contacts to conduct internal research which resulted in completed responses. This often occurred for sites with more complex HVAC systems or sites that used different fuel sources in different areas of their business for either HVAC or domestic hot water.

Where the surveyor provided accurate square footage measurements and partitioning, most data issues could be fixed through either internet research or over the phone. However, for

sites where the square footage measurements were unreliable or the site was known to be partitioned incorrectly, a follow up site visit was generally necessary to correct the issues(though occasionally sites could provide floorplans or mechanical schedules to avoid a second site visit). When follow up appointments were necessary, specific instructions were given to the new surveyor so that they could pinpoint the survey's problems and correct them in a timely fashion. This reduced the amount of time required to complete the follow-up visit. In general, revisits could be completed in less than an hour, with the completed site now passing the QC score threshold.

For a very small number of sites, remediation was not possible. This was generally due to the site going out of business or refusing to answer follow up questions over the phone or on-site. When this occurred, the site was considered incomplete and removed from the data set of completed sites.