# **Data Validation for FY24 Governmentwide Annual Assessment**

The General Services Administration (GSA) created a [R](https://www.r-project.org/) script, outlined below, to systematically validate response data based on specific criteria. These data validation tests primarily employ conditional if-then logic, leveraging the interdependencies between various responses within a reporting entity's submission. Collectively, the script facilitates a dynamic and responsive validation process that identifies potential inconsistencies inherent in the data provided by the entities. Upon detecting such discrepancies, a flag is triggered. While GSA categorized and documented these inconsistencies, no alterations were made to the data during the validation process.

**R Script**

# Load necessary libraries

library(dplyr)

library(stringr)

# Define the file path variable

file\_path <- "YOUR\_FILE\_PATH\_HERE" # Replace with the actual file path

# Read the data from the file path

data <- read.csv(file\_path, header = TRUE, stringsAsFactors = FALSE)

# Remove the second row (duplicate header row)

data <- data[-2, ]

# Identify and rename duplicate column names

names(data)[duplicated(names(data))] <- make.unique(names(data)[duplicated(names(data))])

# Remove columns with blank or empty headings

data <- data[, !sapply(data, function(x) all(is.na(x) | x == ""))]

# Clean the data to remove any unwanted characters or extra spaces

clean\_data <- function(data) {

data <- data %>%

mutate(across(everything(), ~ gsub("[[:space:]]+$", "", .))) %>% # Remove trailing spaces

mutate(across(everything(), ~ gsub("[[:space:]]+", " ", .))) # Remove extra spaces between words

return(data)

}

# Create an empty data frame to store validation results

validation\_results <- data.frame(Check = character(0), Result = character(0), Column = character(0))

# Define a function to add results to the validation data frame

add\_result <- function(check\_name, result, column, message) {

if (nrow(result) > 0) {

result$Validation <- check\_name

result$Column <- column

result$Result <- message

validation\_results <<- rbind(validation\_results, result)

}

}

# Define a function to exclude blanks from the columns involved in a validation

exclude\_blanks <- function(data, columns) {

if (length(columns) == 1) {

# If there's only one column, use a different method to check for blanks

data <- data[!is.na(data[[columns]]) & data[[columns]] != "", ]

} else {

# For multiple columns, use rowSums to filter out rows with any blanks

data <- data[rowSums(is.na(data[, columns]) | data[, columns] == "") == 0, ]

}

return(data)

}

# Convert relevant columns to numeric where necessary

data$Q1\_1\_TEXT <- as.numeric(as.character(data$Q1\_1\_TEXT))

data$Q2 <- as.numeric(as.character(data$Q2))

data$Q25 <- as.numeric(as.character(data$Q25))

data$Q67\_1\_TEXT <- as.numeric(as.character(data$Q67\_1\_TEXT))

data$Q68a\_4\_TEXT <- as.numeric(as.character(data$Q68a\_4\_TEXT))

data$Q68b\_4\_TEXT <- as.numeric(as.character(data$Q68b\_4\_TEXT))

data$Q68c\_4\_TEXT <- as.numeric(as.character(data$Q68c\_4\_TEXT))

data$Q68d\_4\_TEXT <- as.numeric(as.character(data$Q68d\_4\_TEXT))

data$Q69a\_4\_TEXT <- as.numeric(as.character(data$Q69a\_4\_TEXT))

data$Q70\_4\_TEXT <- as.numeric(as.character(data$Q70\_4\_TEXT))

data$Q73a\_1\_TEXT <- as.numeric(as.character(data$Q73a\_1\_TEXT))

data$Q73b\_1\_TEXT <- as.numeric(as.character(data$Q73b\_1\_TEXT))

data$Q73c\_4\_TEXT <- as.numeric(as.character(data$Q73c\_4\_TEXT))

data$Q73d\_4\_TEXT <- as.numeric(as.character(data$Q73d\_4\_TEXT))

data$Q74a\_4\_TEXT <- as.numeric(as.character(data$Q74a\_4\_TEXT))

data$Q75\_1\_TEXT <- as.numeric(as.character(data$Q75\_1\_TEXT))

data$Q92\_1\_TEXT <- as.numeric(as.character(data$Q92\_1\_TEXT))

data$Q93\_1\_TEXT <- as.numeric(as.character(data$Q93\_1\_TEXT))

data$Q96\_1\_TEXT <- as.numeric(as.character(data$Q96\_1\_TEXT))

data$Q97\_1\_TEXT <- as.numeric(as.character(data$Q97\_1\_TEXT))

data$Q98\_1\_TEXT <- as.numeric(as.character(data$Q98\_1\_TEXT))

data$Q99\_1\_TEXT <- as.numeric(as.character(data$Q99\_1\_TEXT))

# Validation 1.1: Federal employee counts - Q1\_1\_TEXT data input should be more than Q2’s data input.

filtered\_data\_1 <- data %>%

filter(!is.na(Q1\_1\_TEXT) & !is.na(Q2))

result\_1 <- filtered\_data\_1 %>%

filter(Q1\_1\_TEXT <= Q2)

add\_result("1.1", result\_1, "Q1\_1\_TEXT vs. Q2", "Q1\_1\_TEXT data input is not greater than Q2 data input.")

# Validation 1.2: Validating Q1\_1\_TEXT and Q2 - Q1\_1\_TEXT and Q2 should not match.

filtered\_data\_2 <- data %>%

filter(!is.na(Q1\_1\_TEXT) & !is.na(Q2))

result\_2 <- filtered\_data\_2 %>%

filter(Q1\_1\_TEXT == Q2)

add\_result("1.2", result\_2, "Q1\_1\_TEXT and Q2", "Q1\_1\_TEXT and Q2 data inputs should not match.")

# Validation 1.3: If Q4 a) is selected, Q30 a) nor b) should be selected.

result\_3 <- data %>%

filter(startsWith(Q4, "a)") & (startsWith(Q30, "a)") | startsWith(Q30, "b)")))

add\_result("1.3", result\_3, "Q30", "Q4 a) is selected but Q30 a) or b) is also selected.")

# Validation 1.4: If Q4 a) is selected, the FTE number in Q2 should be at least 1.

result\_4 <- data %>%

filter(startsWith(Q4, "a)") & Q2 < 1)

add\_result("1.4", result\_4, "Q2", "Q4 a) is selected but Q2 FTE number is less than 1.")

# Validation 1.5: If Q4 b) is selected, FTE count in Q2 should be greater than 0.

filtered\_data\_5 <- exclude\_blanks(data, c("Q4", "Q2"))

result\_5 <- filtered\_data\_5 %>%

filter(startsWith(Q4, "b)") & Q2 <= 0)

add\_result("1.5", result\_5, "Q2", "Q4 b) is selected but Q2 FTE count is 0 or less.")

# Validation 2.1: Conformance test process for web content - If Q9 contains "f)", then, at a minimum, Q10 should be Yes or Q13 should have an answer selected between a) and e).

filtered\_data\_6 <- exclude\_blanks(data, c("Q9", "Q10", "Q13"))

result\_6 <- filtered\_data\_6 %>%

filter(str\_detect(Q9, "f\\)") & !(startsWith(Q10, "a) Yes") | str\_detect(Q13, "^[a-e]\\)")))

add\_result("2.1", result\_6, "Q10/Q13", "Q9 contains 'f)' but Q10 is not 'Yes' and Q13 does not have a) to e) selected.")

# Validation 3.1: If "b) No" is selected in Q10, then Q68c\_4\_TEXT should be 0.

filtered\_data\_7 <- exclude\_blanks(data, c("Q10", "Q68c\_4\_TEXT"))

result\_7 <- filtered\_data\_7 %>%

filter(startsWith(Q10, "b)") & Q68c\_4\_TEXT != 0)

add\_result("3.1", result\_7, "Q68c\_4\_TEXT", "Q10 b) No is selected but Q68c\_4\_TEXT is not 0.")

# Validation 3.2: If "b) No" is selected in Q10, then Q68b\_4\_TEXT should be 0.

filtered\_data\_8 <- exclude\_blanks(data, c("Q10", "Q68b\_4\_TEXT"))

result\_8 <- filtered\_data\_8 %>%

filter(startsWith(Q10, "b)") & Q68b\_4\_TEXT != 0)

add\_result("3.2", result\_8, "Q68b\_4\_TEXT", "Q10 b) No is selected but Q68b\_4\_TEXT is not 0.")

# Validation 3.3: If "b) No" is selected in Q10, then Q73c\_4\_TEXT should be 0.

filtered\_data\_9 <- exclude\_blanks(data, c("Q10", "Q73c\_4\_TEXT"))

result\_9 <- filtered\_data\_9 %>%

filter(startsWith(Q10, "b)") & Q73c\_4\_TEXT != 0)

add\_result("3.3", result\_9, "Q73c\_4\_TEXT", "Q10 b) No is selected but Q73c\_4\_TEXT is not 0.")

# Validation 3.4: If "b) No" is selected in Q10, then Q73b\_1\_TEXT should be 0.

filtered\_data\_10 <- exclude\_blanks(data, c("Q10", "Q73b\_1\_TEXT"))

result\_10 <- filtered\_data\_10 %>%

filter(startsWith(Q10, "b)") & Q73b\_1\_TEXT != 0)

add\_result("3.4", result\_10, "Q73b\_1\_TEXT", "Q10 b) No is selected but Q73b\_1\_TEXT is not 0.")

# Validation 4.1: Manual testing 1 - If Q13 contains "f)", it should report "0" for the number of public internet web pages evaluated by only manual testing (Q68d\_4\_TEXT).

filtered\_data\_11 <- exclude\_blanks(data, c("Q13", "Q68d\_4\_TEXT"))

result\_11 <- filtered\_data\_11 %>%

filter(str\_detect(Q13, "f\\)") & Q68d\_4\_TEXT != 0)

add\_result("4.1", result\_11, "Q68d\_4\_TEXT", "Q13 contains 'f)' but Q68d\_4\_TEXT is not 0.")

# Validation 4.2: Manual testing 2 - If Q13 contains "f)", it should report "0" for the number of public internet web pages evaluated by a combination of both automated and manual testing (Q68b\_4\_TEXT).

filtered\_data\_12 <- exclude\_blanks(data, c("Q13", "Q68b\_4\_TEXT"))

result\_12 <- filtered\_data\_12 %>%

filter(str\_detect(Q13, "f\\)") & Q68b\_4\_TEXT != 0)

add\_result("4.2", result\_12, "Q68b\_4\_TEXT", "Q13 contains 'f)' but Q68b\_4\_TEXT is not 0.")

# Validation 4.3: Manual testing 3 - If Q13 contains "f)", it should report "0" for the number of internal intranet web pages evaluated by manual testing (Q73d\_4\_TEXT).

filtered\_data\_13 <- exclude\_blanks(data, c("Q13", "Q73d\_4\_TEXT"))

result\_13 <- filtered\_data\_13 %>%

filter(str\_detect(Q13, "f\\)") & Q73d\_4\_TEXT != 0)

add\_result("4.3", result\_13, "Q73d\_4\_TEXT", "Q13 contains 'f)' but Q73d\_4\_TEXT is not 0.")

# Validation 4.4: Manual testing 4 - If Q13 contains "f)", it should report "0" for the number of internal intranet web pages evaluated by a combination of both automated and manual testing (Q73b\_1\_TEXT).

filtered\_data\_14 <- exclude\_blanks(data, c("Q13", "Q73b\_1\_TEXT"))

result\_14 <- filtered\_data\_14 %>%

filter(str\_detect(Q13, "f\\)") & Q73b\_1\_TEXT != 0)

add\_result("4.4", result\_14, "Q73b\_1\_TEXT", "Q13 contains 'f)' but Q73b\_1\_TEXT is not 0.")

# Validation 5.1: If "b)" is selected in Q16 or Q17, flag if "a) Yes" is selected in Q21.

filtered\_data\_15 <- exclude\_blanks(data, c("Q16", "Q17", "Q21"))

result\_15 <- filtered\_data\_15 %>%

filter((startsWith(Q16, "b)") | startsWith(Q17, "b)")) & startsWith(Q21, "a) Yes"))

add\_result("5.1", result\_15, "Q21", "Q16 or Q17 b) is selected but Q21 is Yes.")

# Validation 5.2: If "b)" is selected in Q18, flag if "a) Yes" is selected in Q19.

filtered\_data\_16 <- exclude\_blanks(data, c("Q18", "Q19"))

result\_16 <- filtered\_data\_16 %>%

filter(startsWith(Q18, "b)") & startsWith(Q19, "a) Yes"))

add\_result("5.2", result\_16, "Q19", "Q18 b) is selected but Q19 is Yes.")

# Validation 6.1: If Q24 d) is selected, Q51 a), f), or g) should not be selected (and vice versa).

filtered\_data\_17 <- exclude\_blanks(data, c("Q24", "Q51"))

result\_17\_original <- filtered\_data\_17 %>%

filter(str\_detect(Q24, "d\\)") & (startsWith(Q51, "a)") | startsWith(Q51, "f)") | startsWith(Q51, "g)")))

add\_result("6.1", result\_17\_original, "Q24/Q51", "Q24 contains 'd)' but Q51 a), f), or g) is also selected.")

result\_17\_vice\_versa <- filtered\_data\_17 %>%

filter((startsWith(Q51, "a)") | startsWith(Q51, "f)") | startsWith(Q51, "g)")) & str\_detect(Q24, "d\\)"))

add\_result("6.1", result\_17\_vice\_versa, "Q24/Q51", "Q51 a), f), or g) is selected, but Q24 also contains 'd)'.")

# Validation 7.1: Electronic content Conformance Test process - If Q42 b), c), d) or e) is selected, Q9 should contain "b)", "c)", "d)", or "f)" anywhere in the string.

filtered\_data\_19 <- exclude\_blanks(data, c("Q42", "Q9"))

result\_19 <- filtered\_data\_19 %>%

filter((startsWith(Q42, "b)") | startsWith(Q42, "c)") | startsWith(Q42, "d)") | startsWith(Q42, "e)")) &

!str\_detect(Q9, "b\\)|c\\)|d\\)|f\\)"))

add\_result("7.1", result\_19, "Q9", "Q42 b), c), d), or e) is selected, but Q9 does not contain 'b)', 'c)', 'd)', or 'f)'.")

# Validation 7.2: Docs Conformance Test process - If Q55 b), c), d), or e) is selected, Q9 should contain "d)" anywhere in the string.

filtered\_data\_20 <- exclude\_blanks(data, c("Q55", "Q9"))

result\_20 <- filtered\_data\_20 %>%

filter((startsWith(Q55, "b)") | startsWith(Q55, "c)") | startsWith(Q55, "d)") | startsWith(Q55, "e)")) &

!str\_detect(Q9, "d\\)"))

add\_result("7.2", result\_20, "Q9", "Q55 b), c), d), or e) is selected, but Q9 does not contain 'd)'.")

# Validation 7.3: Web Conformance Test Process - If Q52 b) through e) is selected, Q9 should contain "f)" anywhere in the string.

filtered\_data\_21 <- exclude\_blanks(data, c("Q52", "Q9"))

result\_21 <- filtered\_data\_21 %>%

filter((startsWith(Q52, "b)") | startsWith(Q52, "c)") | startsWith(Q52, "d)") | startsWith(Q52, "e)")) &

!str\_detect(Q9, "f\\)"))

add\_result("7.3", result\_21, "Q9", "Q52 b) through e) is selected, but Q9 does not contain 'f)'.")

# Validation 7.4: Manual Web Conformance Test Process - If Q53 b) through e) is selected, Q13 should contain "a)", "b)", "c)", "d)", or "e)" anywhere in the string.

filtered\_data\_22 <- exclude\_blanks(data, c("Q53", "Q13"))

result\_22 <- filtered\_data\_22 %>%

filter((startsWith(Q53, "b)") | startsWith(Q53, "c)") | startsWith(Q53, "d)") | startsWith(Q53, "e)")) &

!str\_detect(Q13, "a\\)|b\\)|c\\)|d\\)|e\\)"))

add\_result("7.4", result\_22, "Q13", "Q53 b) through e) is selected, but Q13 does not contain 'a)', 'b)', 'c)', 'd)', or 'e)'.")

# Validation 7.5: If Q54 b) through e) is selected, Q10 "a) Yes" should be selected.

filtered\_data\_23 <- exclude\_blanks(data, c("Q54", "Q10"))

result\_23 <- filtered\_data\_23 %>%

filter((startsWith(Q54, "b)") | startsWith(Q54, "c)") | startsWith(Q54, "d)") | startsWith(Q54, "e)")) &

!startsWith(Q10, "a) Yes"))

add\_result("7.5", result\_23, "Q10", "Q54 b) through e) is selected but Q10 a) Yes is not selected.")

# Validation 8.1: Sufficiency of acquisition language - Q57 and Q99\_1\_TEXT data points should align with specific conditions.

filtered\_data\_24 <- exclude\_blanks(data, c("Q57", "Q99\_1\_TEXT"))

data$Q57 <- as.numeric(as.character(data$Q57))

data$Q99\_1\_TEXT <- as.numeric(as.character(data$Q99\_1\_TEXT))

result\_24 <- filtered\_data\_24 %>%

filter((Q57 == "a)" & Q99\_1\_TEXT > 0) |

(Q57 == "b)" & Q99\_1\_TEXT > 24.99) |

(Q57 == "c)" & Q99\_1\_TEXT > 59.99) |

(Q57 == "d)" & Q99\_1\_TEXT > 89.99) |

(Q57 == "e)" & Q99\_1\_TEXT < 90))

add\_result("8.1", result\_24, "Q57/Q99\_1\_TEXT", "Q57 and Q99\_1\_TEXT do not align with the expected ranges.")

# Validation 9.1: If Q64 d) or e) are selected, Q25 a) should be selected.

data$Q25 <- as.character(data$Q25)

data$Q64 <- as.character(data$Q64)

filtered\_data\_25 <- exclude\_blanks(data, c("Q64", "Q25"))

result\_25 <- filtered\_data\_25 %>%

filter((startsWith(Q64, "d)") | startsWith(Q64, "e)")) & !startsWith(Q25, "a)"))

add\_result("9.1", result\_25, "Q25", "Q64 d) or e) is selected but Q25 a) is not selected.")

# Validation 10.1: The answer for Q68a\_4\_TEXT should be equal to or less than Q67\_1\_TEXT.

filtered\_data\_26 <- exclude\_blanks(data, c("Q68a\_4\_TEXT", "Q67\_1\_TEXT"))

result\_26 <- filtered\_data\_26 %>%

filter(Q68a\_4\_TEXT > Q67\_1\_TEXT)

add\_result("10.1", result\_26, "Q68a\_4\_TEXT/Q67\_1\_TEXT", "Q68a\_4\_TEXT is greater than Q67\_1\_TEXT.")

# Validation 10.2: Q68b\_4\_TEXT, Q68c\_4\_TEXT, and Q68d\_4\_TEXT should sum to 100%.

filtered\_data\_27 <- exclude\_blanks(data, c("Q68b\_4\_TEXT", "Q68c\_4\_TEXT", "Q68d\_4\_TEXT"))

result\_27 <- filtered\_data\_27 %>%

filter((Q68b\_4\_TEXT + Q68c\_4\_TEXT + Q68d\_4\_TEXT) != 100)

add\_result("10.2", result\_27, "Q68b\_4\_TEXT/Q68c\_4\_TEXT/Q68d\_4\_TEXT", "Q68b\_4\_TEXT, Q68c\_4\_TEXT, and Q68d\_4\_TEXT do not sum to 100%.")

# Validation 11.1: Q69a\_4\_TEXT should be between 0-100.

filtered\_data\_28 <- exclude\_blanks(data, c("Q69a\_4\_TEXT"))

result\_28 <- filtered\_data\_28 %>%

filter(Q69a\_4\_TEXT < 0 | Q69a\_4\_TEXT > 100)

add\_result("11.1", result\_28, "Q69a\_4\_TEXT", "Q69a\_4\_TEXT value is out of range (0-100%).")

# Validation 11.2: Q74a\_4\_TEXT should be between 0-100.

filtered\_data\_29 <- exclude\_blanks(data, c("Q74a\_4\_TEXT"))

result\_29 <- filtered\_data\_29 %>%

filter(Q74a\_4\_TEXT < 0 | Q74a\_4\_TEXT > 100)

add\_result("11.2", result\_29, "Q74a\_4\_TEXT", "Q74a\_4\_TEXT value is out of range (0-100%).")

# Validation 12.1: If Q69a\_4\_TEXT is between 0-90%, then Q70\_4\_TEXT should be greater than 0.

filtered\_data\_30 <- exclude\_blanks(data, c("Q69a\_4\_TEXT", "Q70\_4\_TEXT"))

result\_30 <- filtered\_data\_30 %>%

filter(Q69a\_4\_TEXT < 90 & Q70\_4\_TEXT <= 0)

add\_result("12.1", result\_30, "Q70\_4\_TEXT", "Q69a\_4\_TEXT is between 0-90% but Q70\_4\_TEXT is 0 or less.")

# Validation 13.1: If the percentage of public web pages tested with automated tools (Q68c\_4\_TEXT) is > 75%, Q69a\_4\_TEXT should not indicate 100% conformance.

filtered\_data\_31 <- exclude\_blanks(data, c("Q68c\_4\_TEXT", "Q68a\_4\_TEXT", "Q69a\_4\_TEXT"))

result\_31 <- filtered\_data\_31 %>%

filter((Q68c\_4\_TEXT / Q68a\_4\_TEXT > 0.75) & Q69a\_4\_TEXT == 100)

add\_result("13.1", result\_31, "Q69a\_4\_TEXT", "Q68c\_4\_TEXT percentage exceeds 75%, but Q69a\_4\_TEXT indicates 100% conformance.")

# Validation 13.2: If the percentage of public internet web pages evaluated with automatic testing tools (Q68c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q68a) exceeds 75%, then we can question the extent to which these pages fully conform to Section 508 standards (Q69a).

filtered\_data\_32 <- exclude\_blanks(data, c("Q68c\_4\_TEXT", "Q68a\_4\_TEXT", "Q69a\_4\_TEXT"))

result\_32 <- filtered\_data\_32 %>%

filter((Q68c\_4\_TEXT / Q68a\_4\_TEXT) > 0.75)

add\_result("13.2", result\_32, "Q68c\_4\_TEXT/Q68a\_4\_TEXT/Q69a\_4\_TEXT", "The percentage of public internet web pages evaluated with automatic testing tools (Q68c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q68a) exceeds 75%, which raises questions about the extent to which these pages fully conform to Section 508 standards (Q69a).")

# Validation 14.1: If the percentage of public internet web pages tested with automated tools (Q73c\_4\_TEXT) is > 75%, then Q74a\_4\_TEXT should not indicate 100% conformance.

filtered\_data\_33 <- exclude\_blanks(data, c("Q73c\_4\_TEXT", "Q73a\_1\_TEXT", "Q74a\_4\_TEXT"))

result\_33 <- filtered\_data\_33 %>%

filter((Q73c\_4\_TEXT / Q73a\_1\_TEXT > 0.75) & Q74a\_4\_TEXT == 100)

add\_result("14.1", result\_33, "Q74a\_4\_TEXT", "Q73c\_4\_TEXT percentage exceeds 75%, but Q74a\_4\_TEXT indicates 100% conformance.")

# Validation 14.2: If the percentage of internal intranet web pages evaluated with automatic testing tools (Q73c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q73a) exceeds 75%, then we can question the extent to which these pages fully conform to Section 508 standards (Q74a).

filtered\_data\_34 <- exclude\_blanks(data, c("Q73c\_4\_TEXT", "Q73a\_1\_TEXT", "Q74a\_4\_TEXT"))

result\_34 <- filtered\_data\_34 %>%

filter((Q73c\_4\_TEXT / Q73a\_1\_TEXT) > 0.75)

add\_result("14.2", result\_34, "Q73c\_4\_TEXT/Q73a\_1\_TEXT/Q74a\_4\_TEXT", "The percentage of internal intranet web pages evaluated with automatic testing tools (Q73c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q73a) exceeds 75%, which raises questions about the extent to which these pages fully conform to Section 508 standards (Q74a).")

# Validation 15.1: The answer for Q73a\_1\_TEXT should be equal to or less than Q72\_4\_TEXT.

filtered\_data\_35 <- exclude\_blanks(data, c("Q73a\_1\_TEXT", "Q72\_4\_TEXT"))

result\_35 <- filtered\_data\_35 %>%

filter(Q73a\_1\_TEXT > Q72\_4\_TEXT)

add\_result("15.1", result\_35, "Q73a\_1\_TEXT/Q72\_4\_TEXT", "Q73a\_1\_TEXT is greater than Q72\_4\_TEXT.")

# Validation 15.2: Q73b\_1\_TEXT, Q73c\_4\_TEXT, and Q73d\_4\_TEXT should sum to 100%.

filtered\_data\_36 <- exclude\_blanks(data, c("Q73b\_1\_TEXT", "Q73c\_4\_TEXT", "Q73d\_4\_TEXT"))

result\_36 <- filtered\_data\_36 %>%

filter((Q73b\_1\_TEXT + Q73c\_4\_TEXT + Q73d\_4\_TEXT) != 100)

add\_result("15.2", result\_36, "Q73b\_1\_TEXT/Q73c\_4\_TEXT/Q73d\_4\_TEXT", "Q73b\_1\_TEXT, Q73c\_4\_TEXT, and Q73d\_4\_TEXT do not sum to 100%.")

# Validation 16.1: If Q74a\_4\_TEXT is between 0-90%, then Q75\_1\_TEXT should be greater than 0.

filtered\_data\_37 <- exclude\_blanks(data, c("Q74a\_4\_TEXT", "Q75\_1\_TEXT"))

result\_37 <- filtered\_data\_37 %>%

filter(Q74a\_4\_TEXT < 90 & Q75\_1\_TEXT <= 0)

add\_result("16.1", result\_37, "Q75\_1\_TEXT", "Q74a\_4\_TEXT is between 0-90% but Q75\_1\_TEXT is 0 or less.")

# Validation 17.1: Q93\_1\_TEXT should not exceed Q92\_1\_TEXT.

filtered\_data\_38 <- exclude\_blanks(data, c("Q93\_1\_TEXT", "Q92\_1\_TEXT"))

result\_38 <- filtered\_data\_38 %>%

filter(Q93\_1\_TEXT > Q92\_1\_TEXT)

add\_result("17.1", result\_38, "Q93\_1\_TEXT/Q92\_1\_TEXT", "Q93\_1\_TEXT exceeds Q92\_1\_TEXT.")

# Validation 18.1: Section 508-related complaints numbers 1 - Q97\_1\_TEXT’s numerical input should not exceed Q96\_1\_TEXT’s numerical input.

filtered\_data\_39 <- exclude\_blanks(data, c("Q97\_1\_TEXT", "Q96\_1\_TEXT"))

result\_39 <- filtered\_data\_39 %>%

filter(Q97\_1\_TEXT > Q96\_1\_TEXT)

add\_result("18.1", result\_39, "Q97\_1\_TEXT/Q96\_1\_TEXT", "Q97\_1\_TEXT exceeds Q96\_1\_TEXT.")

# Validation 18.2: Section 508-related complaints numbers 2 - Q98\_1\_TEXT’s numerical input should not exceed Q96\_1\_TEXT’s numerical input.

filtered\_data\_40 <- exclude\_blanks(data, c("Q98\_1\_TEXT", "Q96\_1\_TEXT"))

result\_40 <- filtered\_data\_40 %>%

filter(Q98\_1\_TEXT > Q96\_1\_TEXT)

add\_result("18.2", result\_40, "Q98\_1\_TEXT/Q96\_1\_TEXT", "Q98\_1\_TEXT exceeds Q96\_1\_TEXT.")

# Validation 18.3: If Q36 c), d), or e) is selected, Q94 a) should be selected (and vice versa).

filtered\_data\_18 <- exclude\_blanks(data, c("Q36", "Q94"))

result\_18\_original <- filtered\_data\_18 %>%

filter((startsWith(Q36, "c)") | startsWith(Q36, "d)") | startsWith(Q36, "e)")) & !startsWith(Q94, "a)"))

add\_result("18.3", result\_18\_original, "Q36/Q94", "Q36 c), d), or e) is selected, but Q94 a) is not selected.")

result\_18\_vice\_versa <- filtered\_data\_18 %>%

filter(startsWith(Q94, "a)") & !(startsWith(Q36, "c)") | startsWith(Q36, "d)") | startsWith(Q36, "e)")))

add\_result("18.3", result\_18\_vice\_versa, "Q36/Q94", "Q94 a) is selected, but Q36 c), d), or e) is not selected.")

# Save the results

if (nrow(validation\_results) > 0) {

write.csv(validation\_results, "FY24\_validation\_results.csv", row.names = FALSE)

}

Table 1 provides an overview of data validation tests performed by the R script. Each test is uniquely identified by a number consisting of two parts: the whole number and the decimal. The whole number denotes the general topic or group of tests, and the decimal distinguishes individual conditions (or subtopic) within the general topic. Each check is also accompanied by a descriptive title. The "Validation Rationale" column explains the purpose of each validation test, detailing the specific conditions and criteria being examined within the reporting entity’s data. The "Count" column indicates the number of validation failures per validation test. Lastly, “Implications” details the possible impact of the validation failures.

**Table 1. Summary of validation tests.**

| **ID** | **Topic** | **Subtopic** | **Validation Rationale** | **Total Count** | **Implications** |
| --- | --- | --- | --- | --- | --- |
| 1.1 | Federal employee counts and 508 PM status | Federal employee counts | Q1 data input should be more than Q2’s data input. | 14 | If the total for Q1 is less than or equal to Q2, this means entities misunderstood Q1 and misreported their overall entity federal staff count. |
| 1.2 | Federal employee counts and Section 508 PM status | Validating Q1 and Q2 | Q1 and Q2 should not match | 8 | If Q1 and Q2 have the same number, entities misunderstood the differences between each Question. This means that this data is likely unreliable. |
| 1.3 | Federal employee counts and Section 508 PM status | Section 508 PM status | If Q4 a) is selected, Q30 a) nor b) should be selected. | 0 | A mismatch would suggest either the Section 508 PM does not work full-time, or the position does not have clear roles and responsibilities. |
| 1.4 | Federal employee counts and Section 508 PM status | Section 508 PM status and FTE 1 | If Q4 a) is selected, the FTE number in Q2 should be at least 1. | 5 | A mismatch would suggest either the Section 508 PM is not full-time, or the position lacks clear roles and responsibilities. |
| 1.5 | Federal employee counts and Section 508 PM status | Section 508 PM status and FTE 2 | If Q4 b) is selected, FTE count in Q2 should be greater than 0. | 6 | The Section 508 Program Manager is either not a federal employee or there is a number misreported either in the time spent by the Section 508 Program Manager or the number of FTEs. Entities may have underreported federal FTEs, likely due to the limited number of hours spent by the Section 508 PM per week and thus the entity selected 0 instead of a fraction less than 1, or overreported having a part-time Section 508 PM. |
| 2.1 | Conformance test process for web content | Conformance test process for web content | If Q9 f) is selected, then, at a minimum, Q10 should be Yes or Q13 should have an answer selected between a) and e). | 14 | To have a Section 508 conformance test process for web content, entities must be using an automated tool, manual inspection, or both. If Q9 f) is selected but no automated or manual testing methodologies are noted in response to Q10, follow-up is needed to understand how entities perform testing. |
| 3.1 | Automated testing tools | Automated testing tools 1 | If an entity selects "b) No" in Q10 to indicate it does not use an automated accessibility testing tool, it should report "0" for the number of public internet web pages evaluated exclusively through automated testing (Q68c). | 13 | Entities may have misunderstood the testing tool referenced in Q10 and Q68c, leading to lower confidence in results. Entities may utilize another entity to perform public internet testing such as a parent agency that they did not account for when answering. |
| 3.2 | Automated testing tools | Automated testing tools 2 | If an entity selects "b) No" in Q10 to indicate it does not use an automated accessibility testing tool, it should report "0" for the number of public internet web pages evaluated through both automated and manual testing (Q68b). | 23 | Entities may have misunderstood the testing tool referenced in Q10 and Q68b or may utilize semi-automated developer tools that combine both automated tests a user must run manually, coupled with manual testing (Q68d). Entities may utilize another entity to perform public internet testing such as a parent agency that they did not account for when answering. |
| 3.3 | Automated testing tools | Automated testing tools 3 | If an entity selects "b) No" in Q10 to indicate it does not use an automated accessibility testing tool, it should report "0" for the number of internal intranet web pages evaluated exclusively through automated testing (Q73c). | 3 | Entities may have misunderstood the testing tool referenced in Q10 and Q73c, leading to lower confidence in results. Entities may utilize another entity to perform public internet testing such as a parent agency that they did not account for when answering. |
| 3.4 | Automated testing tools | Automated testing tools 4 | If an entity selects "b) No" in Q10 to indicate it does not use an automated accessibility testing tool, it should report "0" for the number of internal intranet web pages evaluated through both automated and manual testing (Q73b). | 8 | Entities may have misunderstood the testing tool referenced in Q10 and Q73b or may utilize semi-automated developer tools that combine both automated tests a user must run manually, coupled with manual testing (Q73d). Entities may utilize another entity to perform public internet testing such as a parent agency that they did not account for when answering. |
| 4.1 | Manual testing | Manual testing 1 | If an entity indicates it does not use a manual or hybrid testing methodology by selecting "f) N/A - entity does not use a manual and/or hybrid testing methodology" in Q13, then it should report "0" for the number of public internet web pages evaluated by only manual testing (Q68d). | 5 | Entities may have misunderstood the testing tool referenced in Q13 and Q68d, possibly leading to overreporting the number of pages for Q68d, performing manual testing without a manual or hybrid testing methodology, or not reporting the presence of a manual or hybrid testing methodology for Q13 despite having one. This results in low confidence in the data reported for both criteria. |
| 4.2 | Manual testing | Manual testing 2 | If an entity indicates that it does not use a manual or hybrid testing methodology by selecting "f) N/A - entity does not use a manual and/or hybrid testing methodology" in Q13, then it should report "0" for the number of public internet web pages evaluated by a combination of both automated and manual testing (Q68b). | 6 | Entities may have misunderstood the testing tool referenced in Q13 and Q68b, leading possibly to overreporting the number of pages for Q68b, performing manual testing without a manual or hybrid testing methodology, or not reporting the presence of a manual or hybrid testing methodology for Q13 despite having one. This results in low confidence in the data reported for both criteria. |
| 4.3 | Manual testing | Manual testing 3 | If an entity indicates it does not use a manual or hybrid testing methodology by selecting "f) N/A - entity does not use a manual and/or hybrid testing methodology" in Q13, then it should report "0" for the number of internal intranet web pages evaluated by manual testing (Q73d). | 2 | Entities may have misunderstood the testing tool referenced in Q13 and Q73d, possibly leading to overreporting the number of pages for Q73d, performing manual testing without a manual or hybrid testing methodology, or not reporting the presence of a manual or hybrid testing methodology for Q13 despite having one. This results in low confidence in the data reported for both criteria. |
| 4.4 | Manual testing | Manual testing 4 | If an entity indicates it does not use a manual or hybrid testing methodology by selecting "f) N/A - entity does not use a manual and/or hybrid testing methodology" in Q13, then it should report "0" for the number of internal intranet web pages evaluated by a combination of both automated and manual testing (Q73b). | 1 | Entities may have misunderstood the testing tool referenced in Q13 and Q73b, leading possibly to overreporting the number of pages for Q73b, performing manual testing without a manual and/or hybrid testing methodology, or not reporting the presence of a manual or hybrid testing methodology for Q13 despite having one. This results in low confidence in the data reported for both criteria. |
| 5.1 | Exceptions | Exceptions 1 | If b) is selected for Q16 or Q17, flag if “Yes” is selected for Q21. | 18 | A defined process for reviewing and approving exceptions implies a system for tracking them. If an entity has a process but does not know how many exceptions were approved, this may suggest an ineffective process. |
| 5.2 | Exceptions | Exceptions 2 | If b) is selected for Q18, flag if “Yes” is selected for Q19. | 14 | A defined process for reviewing and approving exceptions implies a system for tracking them. If an entity has a process but does not know how many exceptions were approved, this may suggest an ineffective process. |
| 6.1 | User testing with PWDs | Testing with PWDs | If Q24 d) is selected, Q51 a), f), or g) should not be selected (and vice versa). | 56 | Since both questions relate to conducting user testing with PWD, there should be parity in the responses. A mismatch suggests that an entity may aspire to include PWDs but in practice are not. |
| 7.1 | Content conformance test processes | Electronic content conformance test process | If Q42 b), c), d) or e) is selected, Q9 should have at least one of the following selected: b), c), d) or f). | 18 | For an entity to be able to test electronic content prior to publication, they should also be performing the evaluation with a test process. If no test process for electronic content is selected in response to Q9, analysis is needed further to understand how entities are testing content. |
| 7.2 | Content conformance test processes | Document conformance test process | If Q55 b), c), d) or e) is selected, Q9 d) should be selected. | 48 | For an entity to be able to test electronic documents, they should also be performing the evaluation with a test process. If no test process for electronic documents is selected in response to Q9, analysis is needed further to understand how entities are testing content. |
| 7.3 | Content conformance test processes | Web conformance test Process | If Q52 b) through e) is selected, Q9 f) should also be selected. | 21 | To produce reliable web content test results using standardized processes for Section 508, entities must be using a test process. If Q9 f) is not selected, follow-up is needed to understand how entities perform testing. |
| 7.4 | Content conformance test processes | Manual web conformance test process | If Q53 b) through e) is selected, at least one selection in Q13 a) to e) should also be selected. | 18 | To complete comprehensive manual Section 508 testing, entities must use a test process. If none of Q13 a) through e) is selected, follow-up is needed to understand how entities are performing testing. |
| 7.5 | Content conformance test processes | Automated web conformance test process | If Q54 b) through e) is selected, Q10 a) should be selected. | 40 | To complete comprehensive automated Section 508 testing, entities must use a test tool. If Q10 a) is not selected, follow-up is needed to understand how entities are performing testing. There also may have been a misunderstanding with Q10’s language or phrasing. |
| 8.1 | Acquisition language | Sufficiency of acquisition language | Q57 and Q99 data points should be comparable. Outliers should not occur. | 0 | Both questions address sufficiency of Section 508 language in solicitations. A mismatch may suggest agencies misrepresented the actual number in Q57, or the specific solicitations reviewed for Q99 were not an average sample for the entity. A mismatch may suggest the entity has room for improvement to ensure standard processes are followed for each solicitation. |
| 9.1 | Mandatory training | Mandatory training | If Q64 d) or e) are selected, Q25 a) should be selected. | 0 | If there is mandatory Section 508 awareness training (Q64), there should be a required mandatory training for everyone (Q25). A mismatch may occur because entities did not understand the question. |
| 10.1 | General checks for public web pages | Number of public internet web pages | The answer for Q68a should be equal to or less than Q67. | 3 | If 68a is larger than Q67, this suggests some pages were tested more than once. |
| 10.2 | General checks for public web pages | Test method to evaluate public internet web pages | Q68b, Q68c, and Q68d should equal 100%. | 40 | If the sum of these numbers is less than 100%, the reporting entity possibly did not know the testing methods employed. If the sum of these numbers is more than 100%, it suggests that there was a misunderstanding with terminology and data was not entered correctly in the respective answer selections. The implications are that there may be some question as to whether comprehensive testing was performed and impact the accuracy of Q69a. |
| 11.1 | Valid percentage range for conformance | Valid data input range for public internet web pages | Q69a, answer selection a) should be between 0 and 100 due to being a percentage. | 0 | A number above 100 is an invalid response due to the question asking for a percentage (out of 100). |
| 11.2 | Valid percentage range for conformance | Valid data input range for internal intranet web pages | Q74a, answer selection a) should be between 0 and 100 due to being a percentage. | 0 | A number above 100 is an invalid response due to the question asking for a percentage (out of 100). |
| 12.1 | Non-conformance issues for public internet web pages | Average non-conformance issues for public internet web pages | If Q69a a) has an answer selection between 0-90%, Q70 a) should have an answer above 0 noted in a). Unknowns selected in Q70 b) should be omitted from this test. | 6 | If all web pages do not conform in Q69 a), then there should be an average of at least one defect on each page. If not, it is possible that the reporting entity misreported or the defects, on average, were calculated to a miniscule amount that rounded to 0. |
| 13.1 | Public internet web pages evaluated with automated testing tools | Percentage of public internet web pages tested with automated tools 1 | If the percentage of public internet web pages evaluated with automatic testing tools (Q68c) compared to the total number of public internet web pages evaluated for Section 508 conformance (Q68a) exceeds 75%, then 100% of these pages should not fully conform to Section 508 standards (Q69a). Some aspects of web conformance can only be evaluated manually. | 4 | Entities who relied heavily on using only automated accessibility testing tools may have overinflated conformance to Section 508 web requirements; automated tools do not test for every Section 508 standard and thus roughly 70% of Section 508 web standards were not tested using only automated tools. |
| 13.2 | Public internet web pages evaluated with automated testing tools | Percentage of public internet web pages tested with automated tools 2 | If the percentage of public internet web pages evaluated with automatic testing tools (Q68c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q68a) exceeds 75%, then we can question the extent to which these pages fully conform to Section 508 standards (Q69a). Some aspects of web conformance can only be evaluated manually. | 17 | Entities who relied heavily on using only automated accessibility testing tools may have overinflated conformance to Section 508 web requirements; automated tools do not test for every Section 508 standard and thus roughly 70% of Section 508 web standards were not tested using only automated tools. |
| 14.1 | Internal intranet web pages evaluated with automated tools | Percentage of internal intranet tested web pages evaluated with automated tools 1 | If the percentage of internal intranet web pages evaluated with automatic testing tools (Q73c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q73a) exceeds 75%, then 100% of these pages should not fully conform to Section 508 standards (Q74a). Some aspects of web conformance can only be evaluated manually. | 0 | Entities who relied heavily on using only automated accessibility testing tools may have overinflated conformance to Section 508 web requirements; automated tools do not test for every Section 508 standard and thus roughly 70% of Section 508 web standards were not tested using only automated tools. |
| 14.2 | Internal intranet web pages evaluated with automated testing tools | Percentage of internal intranet tested web pages evaluated with automated tools 2 | If the percentage of internal intranet web pages evaluated with automatic testing tools (Q73c) compared to the total number of internal intranet web pages evaluated for Section 508 conformance (Q73a) exceeds 75%, then we can question the extent to which these pages fully conform to Section 508 standards (Q74a). Some aspects of web conformance can only be evaluated manually. | 20 | Entities who relied heavily on using only automated accessibility testing tools may have overinflated conformance to Section 508 web requirements; automated tools do not test for every Section 508 standard and thus roughly 70% of Section 508 web standards were not tested using only automated tools. |
| 15.1 | General checks for internal intranet web pages | Number of internal intranet web pages | The answer for Q73a should be equal to or less than Q72. | 22 | If Q73a is larger than Q72, this suggests some pages were tested more than once. |
| 15.2 | General checks for internal intranet web pages | Test method of internal intranet web pages | Q73b, Q73c, and Q73d should equal 100%. | 23 | If the sum of these numbers is less than 100%, the reporting entity possibly did not know the testing methods employed. If the sum of these numbers is more than 100%, it suggests that there was a misunderstanding with terminology and data was not entered correctly in the respective answer selections. The implications are that there may be some question as to whether comprehensive testing was performed and impact the accuracy of Q73a. |
| 16.1 | Non-conformance issues for internal intranet web pages | Average non-conformance issues for internal intranet web pages | If Q74a a) has an answer selection between 0-90%, Q75 a) should have an answer above 0 noted in a). Unknowns selected in Q75 b) should be omitted from this test. | 23 | If all web pages do not conform in Q74a, then there should be an average of at least one defect on each page. If not, it is possible that the reporting entity misreported or the defects, on average, were calculated to a miniscule amount that rounded to 0. |
| 17.1 | Public feedback responses | Public feedback responses | Q93’s numerical input should not exceed Q92’s numerical input. | 1 | If Q93’s numerical input exceeds Q92’s numerical input, it is possible the reporting entity incorrectly included resolution of feedback outside of the reporting period or miscalculated the total in Q92. |
| 18.1 | Complaints | Section 508-related complaints numbers 1 | Q97’s numerical input should not exceed Q96’s numerical input. | 0 | If Q97’s numerical input exceeds Q96’s numerical input, it is possible the reporting entity incorrectly included resolution of complaints outside of the reporting period or miscalculated the total in Q96. |
| 18.2 | Complaints | Section 508-related complaints numbers 2 | Q98’s numerical input should not exceed Q96’s numerical input. | 0 | If Q98’s numerical input exceeds Q96’s numerical input, it is possible the reporting entity incorrectly included resolution of complaints outside of the reporting period or miscalculated the total in Q96. |
| 18.3 | Complaints | Section 508 complaint process/procedure | If Q36 c), d), or e) is selected, Q94 a) should be selected (and vice versa). | 56 | Since both questions ask about a documented process or procedure to Section 508 complaints, there should be parity in the answers. A mismatch suggests a misunderstanding of the question. |