Cálculo de la V de Aiken con R

Alfonso Cano Robles

1 V de Aiken

El objetivo de este documento es presentar el cálculo de la V de Aiken y su intervalo de confianza con el software `R`, su interpretación y formas de presentación en la la interfaz gráfica de usuario (GUI) `RKWard`. Esto se debe a que no se ha encontrado en la literatura especializada un documento que sintetice estos elementos y los presente de manera coherente. Para lograr lo anterior, se inicia con una definición de V. Posteriormente, se presenta el intervalo de confianza junto con sus principales ventajas. En seguida se describe su cálculo en `R` a través de la propuesta de construcción de una función. Luego, se presenta el uso de la propuesta a través de la presentación de un caso concreto. En seguida, se muestra la interpretación de los resultados y de su interpretación, además se propone un gráfico de medias con barras de error como auxiliar en la discriminación de los resultados. Finalmente, se presenta una manera de visualizar los resultados.

1.1 Definición

El coeficiente de la V de Aiken (1980), asume que cada uno de los N evaluadores del cuestionario examina todos los reactivos que lo componen e indica, a través de la elección de un valor numérico en una escala de medición intervalar de carácter discreto (números naturales). En este sentido el o la evaluador expresa su juicio a través de la selección de una categoría ordinal ("c-category"), el número natural, sobre la validez del contenido de cada ítem en el cuestionario. Cuando el total de jueces (N) ha emitido sus deliberaciones, se asigna el valor de 0 para cada n_0, que se encuentre en la categoría ordinal más baja, por ello se asigna el peso de 1 a la n_1 en la siguiente categoría superior, y así para cada n_{c-1} de las evaluaciones en las categorías más altas (c.°). Entonces Aiken (1980, 1985), lo define como:

$$V = \sum_{i=1}^{c-1} \frac{in_i}{N(c-1)}$$

Para este ejercicio se toma la fórmula que nos aportan Penfield y Giacobbi (2004).

$$V = \frac{\bar{X} - l}{k}$$

Con el objetivo de resumir la magnitud de las calificaciones obtenidas por de los expertos, y para hacer un test de hipótesis específicas con respecto a los valores obtenidos para la población. En este caso $\bar{X}$ representa la media de las calificaciones de los jueces. Y k representa el rango de valores posibles en la escala de calificación.

Para realizar la interpretación de la V de Aiken, Oksaviona,Islami y Nasir (2023), retoman la propuesta criterios de rangos de Azwar (2012) como:

Y aunque la V de Aiken provee un marco útil para hacer interpretaciones sobre el nivel de relevancia de contenido de un reactivo (ítem), el proceso inferencial para hacer conclusiones tiene desventajas. Primero, es que el valor crítico de V (V_p) es 0.5, un valor que resulta arbitrario, debido a que usar valores más altos pudiera ser de mayor interés para incluirlos en la escala. En segundo lugar, el cálculo binomial para hacer la prueba de hipótesis sobre la cola de probabilidad puede ser intensivo, por lo que se debe realizar a mano o con la ayuda de software. En tercero, la naturaleza discreta de los datos presenta dificultades para hacer

Cuadro 1: Interpretación de la V de Aiken

Score.Average.Interval	Criteria
$0.80 < V \leq 1.00$	Very High
$0.60 < V \leq 0.80$	High
$0.40 < V \leq 0.60$	Enough
$0.20 < V \leq 0.40$	Low
$0.00 < V \leq 0.20$	Very Low

[a] Oksaviona, Islami y Nasir con base en Azwar (2012)

conclusiones inferenciales cuando el número de evaluadores es pequeño, debido a que no corresponden de manera precisa con el error Tipo I, por lo que las conclusiones pueden ser inexactas. Cuarto, el resultado dice poco sobre la el valor de V_p. En quinto lugar, la prueba de hipótesis no aporta información sobre el error de V como estimación de V_p.

2 Intervalo de confianza para la V de Aiken

La construcción de un intervalo de confianza para V_p radica en que V no se encuentra normalmente distribuida. Las Propiedades del intervalo de confianza es que sea asimétrica y no dependa de una distribución normal en la proporción de la muestra. Se requiere que sea efectiva y precisa, cuando el tamaño de la muestra es pequeño y la proporción de la población es extrema, como muestra Penfield y Giacobbi (2004) con base en Newcombe (1998)) y Wilson (1927).

Penfield y Giacobbi (2004), propone que utilizar marcador del intervalo de confianza como un método para construir el intervalo de confianza para V_p que considera como una mejora en los métodos inferenciales disponibles para la interpretación de la evaluación de la relevancia de contenido de un ítem.

El límite inferior (L) y el mayor (U) se pueden calcular con un valor $C\%$ para un V_p que se puede obtener la siguiente formula desarrollada originalmente por Wilson (1927):

$$L = \frac{2nkV + z^2 - z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)}$$

Y de manera complementaria, para el límite superior:

$$U = \frac{2nkV + z^2 + z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)}$$

En las dos ecuaciones anteriores, z corresponde a un valor de la distribución normal tal como $C\%$ en donde el área de de la distribución se encuentra entre $-z$ y z (Por ejemplo, para un 95% se utiliza un intervalo de confianza de 1.96 y para el 90% el de 1.645). La simplificación de la derivación de las ecuaciones para el límite inferior y superior se presentan como:

$$L = \frac{A - B}{C}$$

Y por

$$U = \frac{A + B}{C}$$

Dónde

$$A = 2nkV + z^2$$

$$B = z\sqrt{4nkV(1-V) + z^2}$$

$$C = 2(nk + z^2)$$

La calificación del intervalo de confianza tiene la propiedad deseable de ue sea asimétrico sobbre V. Si V es mayor que 0.5, entonces el intervalo de confianza se extenderá más por debajo de V que por encima de V. Y su es menor que 0.5, entonces el valor del intervalo se extenderá más por encima de V que por debajo de V. Adicionalmente, los límites no se pueden extender por debajo de 0 o por encima de 1.0, y resuelve un problema de los límites de intervalo de confianza cuando se aplica el intervalo de Wald a variables limitadas. (Penfield y Giacobbi 2004)

A diferencia de la prueba de hipótesis propuesta por Aiken (1980, 1985), el uso de intervalos de confianza permite permite a los investigadores establecer criterios de revisión de ítems más estrictos o flexibles según sus necesidades. Por ejemplo, puede utilizar la hipótesis nula de que $V_p = 0.75$, lo que se asocia con una calificación promedio de 4 en una escala de 5 puntos. Aunque el criterio de 0.75 puede ser demasiado estricto en la práctica, y sugieren que los investigadores pueden optar por valores de criterio más flexibles (p. ej., V0 = 0,4) o tasas de error de tipo I más altas (p. ej., $\alpha = 0.10$), especialmente cuando el número de evaluadores expertos es reducido.

3 Cálculo de la V de Aiken con R

En el Bloque siguiente, se crea en `R` el objeto de función que permite calcular la V de Aiken general con medias de todos Ítems que se presenten en el set de datos y cada un o de los límites correspondientes de su Intervalo de Confianza (CI). Es decir, calcula el valor de V para cada Ítem y el limite Inferior (I) y superior (U) de su IC.

```
v_aiken <- function(
    x, # Es un data frame donde cada fila es un ítem y cada columna contiene las calificaciones que cada evaluador asignó.
    lo, # Es el valor mínimo (lowest) posible en la escala.
    hi, # Es el valor máximo (highest) posible en la escala.
    p) { # Es la proporción del nivel de confianza.
  n <- ncol(x) # Devuelve el número de columnas, es decir, el número de evaluadores.
  i <- nrow(x) # Devuelve el número de filas y representa el número de ítems.
  k <- (hi - lo) # Es la distancia desde "lo" hasta "hi". Es el rango de posibles elecciones discretas.
  z <- qnorm((1 - (p)) / 2, mean = 0, sd = 1, lower.tail = FALSE) # Encuentra el valor z (en unidades de desviación estándar) que corresponde a una probabilidad acumulada dada.
  S <- rowSums(x - lo) # "S" resta elemento por elemento el valor de lo y suma los valores resultantes en cada fila.
  V <- S / (n * k) # "Es el valor de la V de Aiken calculado por la fórmula.
  # Cálculo del IC por medio de la simplificación de la derivación de las ecuaciones para
  el límite inferior y superior
  A <- (2 * n * k * V) + (z^2) # Cálculo de A
  B <- (z * (sqrt(4 * n * k * V * (1 - V) + (z^2)))) # Cálculo de B
  C <- (2 * ((n * k) + (z^2))) # Cálculo de C
  # Cálculo del límite del intervalo de confianza inferior (Lower)
  L <- (A - B) / C
  # Cálculo del límite del intervalo de confianza superior (Upper).
```

```r
  U <- (A + B) / C
  # Crear el data frame "df".
  df <- data.frame(
    cbind(
      "V" =  V,
      "CI_L" = L,
      "CI_U" = U
    )
  )
  # Crear la tabla con nombres con  "Ítem_#".
  rownames(df) <- paste0("Ítem_", 1:i) # Crea el nombre de la fila correspondiente al
  número de cada "Ítem" como se encuentra en el marco de datos x.
  means_list <- list() # Crea la lista llamada "means_list".
  for (col_name in names(df)) {
    # Aplicar la función mean a la columna actual.
    mean_value <- mean(df[[col_name]])
    # Almacenar el resultado en la lista con el nombre de la columna.
    means_list[[col_name]] <- mean_value
  }
  # Crear un data frame llamado "Medias" a partir de la lista "means_list".
  means_df <- data.frame(Medias = unlist(means_list))
  v_list <- list()
  v_list[["v_ci"]] <- df
  v_list[["means_v"]] <- means_df
  # Crea una lista vacía llamada "parameters"
  parameters <- list()
  # Definir los nombres de los objetos a copiar.
  noms_par <- c("n", "k", "p", "z", "lo", "hi", "i")
  # Loop para copiar los objetos a la lista "parameters".
  for (e in 1:length(noms_par)) {
    nombre_actual <- noms_par[e]
    parameters[[nombre_actual]] <- get(nombre_actual)
  }
  v_list[["parameters"]] <- parameters
  return(v_list)
}
```

Por lo anterior, los argumentos que requiere esta función son los siguientes:

```r
v_aiken(x, lo, hi, p)
```

- `x` es un *data.frame* donde cada fila representa las evaluaciones para cada ítem y cada columna contiene las calificaciones que cada evaluador asignó.
- `lo` es el valor mínimo posible en la escala.
- `hi` es el Valor máximo posible en la escala.
- `p` es la proporción del valor del nivel confianza, es decir la probabilidad acumulada dada. Los valores que más se utilizan son 0.90, 0.95 y 0.99.

Para encontrar el valor z (en unidades de desviación estándar) que corresponde a una probabilidad acumulada dada se emplea la función `qnorm()`.

```r
qnorm((1 - (p)) / 2, mean = 0, sd = 1, lower.tail = FALSE)
```

Cuadro 2: Tabla de evaluación por siete Jueces (r) a veinte Ítems

	r1	r2	r3	r4	r5	r6	r7
1	2	5	5	5	5	5	5
2	2	2	3	3	3	4	4
3	2	3	4	4	4	5	5
4	2	2	3	4	5	5	5
5	3	3	4	5	5	5	5
6	3	4	4	5	5	5	5
7	3	3	4	4	4	4	5
8	1	2	3	3	4	5	5
9	1	2	3	3	3	4	5
10	1	3	5	5	5	5	5
11	4	4	4	4	4	5	5
12	4	4	5	5	5	5	5
13	2	3	4	4	4	5	5
14	1	2	3	4	4	4	4
15	4	5	5	5	5	5	5
16	4	4	5	5	5	5	5
17	3	4	4	4	4	5	5
18	3	3	3	4	4	4	4
19	3	4	4	5	5	5	5
20	3	4	4	4	5	5	5

[a] Elaboración propia con base en Penfield y Giacobbi (2004)

3.1 Cálculo de la V de Aiken

Para el siguiente ejercicio, se han retomado los datos de Penfield y Giacobbi(2004), y se han construido en la forma que requiere `x` y se le ha nombrado `test_v`, como se muestra en el el siguiente bloque de texto. La tabla con estos datos se puede observar en el Cuadro 2.

```
test_v <- data.frame(cbind(
  "r1" = c(2, 2, 2, 2, 3, 3, 3, 1, 1, 1, 4, 4, 2, 1, 4, 4, 3, 3, 3, 3),
  "r2" = c(5, 2, 3, 2, 3, 4, 3, 2, 2, 3, 4, 4, 3, 2, 5, 4, 4, 3, 4, 4),
  "r3" = c(5, 3, 4, 3, 4, 4, 4, 3, 3, 5, 4, 5, 4, 3, 5, 5, 4, 3, 4, 4),
  "r4" = c(5, 3, 4, 4, 5, 5, 4, 3, 3, 5, 4, 5, 4, 4, 5, 5, 4, 4, 5, 4),
  "r5" = c(5, 3, 4, 5, 5, 5, 4, 4, 3, 5, 4, 5, 4, 4, 5, 5, 4, 4, 5, 5),
  "r6" = c(5, 4, 5, 5, 5, 5, 4, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5),
  "r7" = c(5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5)
))
```

En este caso la categoría mínima es 1 (`lo = 1`), la máxima es 5 (`hi = 5`) y la proporción de valor del nivel de confianza deseado es 0.95 (`p = .95`), es decir un nivel de significancia de $\alpha = 0.05$. En el Código siguiente:

```
aiken_vals <- v_aiken(
  x = test_v,
  lo = 1,
  hi = 5,
  p = .95
)
```

Como se mencionó más arriba se puede también utilizar un $\alpha = 0.10$, si se coloca una probabilidad acumulada

Cuadro 3: Medias de evaluación de la calificación para todo el instrumento por los siete jueces

	Medias
V	0.752
CI_L	0.578
CI_U	0.865

Cuadro 4: Tabla de evaluación por siete jueces3

	V	CI_L	CI_U
Ítem_1	0.893	0.728	0.963
Ítem_2	0.500	0.326	0.674
Ítem_3	0.714	0.529	0.847
Ítem_4	0.679	0.493	0.821
Ítem_5	0.821	0.644	0.921
Ítem_6	0.857	0.685	0.943
Ítem_7	0.714	0.529	0.847
Ítem_8	0.571	0.391	0.735
Ítem_9	0.500	0.326	0.674
Ítem_10	0.786	0.605	0.898
Ítem_11	0.821	0.644	0.921
Ítem_12	0.929	0.774	0.980
Ítem_13	0.714	0.529	0.847
Ítem_14	0.536	0.358	0.705
Ítem_15	0.964	0.823	0.994
Ítem_16	0.929	0.774	0.980
Ítem_17	0.786	0.605	0.898
Ítem_18	0.643	0.458	0.793
Ítem_19	0.857	0.685	0.943
Ítem_20	0.821	0.644	0.921

de “0.90”.

El objeto que se obtiene, en este caso es `aiken_vals`, designado arriba, es una lista que contiene los objetos `means_v`, `parameters` y `v_ci`. El primero contiene la columna `Medias`con `V`, `CI_L` y `CI_L`, que corresponden al cálculo de la media para todo el set de datos proporcionado para la V de Aiken, y los límites inferior (L) y superior (U) del intervalo de confianza (CI). El segundo objeto, contiene los parámetros utilizados en la fórmula: `n` para el número de evaluadores; `k` para el número de clases, `p` el nivel de confianza expresado en proporción; `z` el valor en unidades de desviaciones estándar que corresponde a `p`; `lo` el valor mínimo que puede aparecer en la escala; `hi` corresponde al valor más alto posible; finalmente, `i` corresponde al número de ítems.

3.2 Los resultados

En el Cuadro 3, se muestran las medias de V, el límite inferior y el límite superior para todo el set de datos calculados con $p = .95$.

Por otro lado, en el Cuadro 4, se muestran las medias para cada ítem en particular.

En el gráfico de la Figura 1, se muestran de barras para la media de evaluaciones obtenidas para cada ítem, con barras de error para cada intervalo de confianza. También, se presenta una linea punteada en el eje horizontal, para marcar el criterio por el cual se pueden poner en revisión los ítems cuyo límite inferior sea menor al criterio seleccionado, en este caso $CI_L < 0.5$.

```
# Preparar
library("ggplot2")
library("tibble")
# Calcular
g <- aiken_vals[["v_ci"]] %>%
  rownames_to_column(var = "Items") %>%
  ggplot(aes(
    x = reorder(Items, V, decreasing = FALSE),
    y = V,
    ymin = CI_L,
    ymax = CI_U
  )) +
  geom_col(fill = "lightblue") +
  geom_errorbar(width = 0.5) +
  ylim(0, 1) +
  geom_hline(
    yintercept = 0.5,
    linetype = "dashed",
    color = "red"
  ) +
  theme(
    plot.background = element_rect(fill = "transparent", color = NA)
  ) +
  coord_flip() +
  ylab("V") +
  xlab("Ítem") +
  labs(
    title = "Gráfico de V de Aiken por Ítem",
    subtitle = "CI en barras de error con p = .95"
  )
# Imprimir
print(g)
```

En el caso de la figura 1, se han ordenado el eje de los Ítems por la media obtenida de `V` y se puede observar que los ítems 4, 18, 14, 9 y 2, requieren ser revisados si se sigue el criterio de mencionado más arriba.

3.3 Presentar en RKWard

Presentar la media general y la tabla por cada Ítem, con el siguiente código, para obtener la salida en HTML.

```
local({
  rk.header(
    title = "V de Aiken", # Asigna el texto contenido en el título.
    parameters = aiken_vals[["parameters"]]
  ) # Presenta los elementos contenidos en la lista "parameters".
  rk.results(aiken_vals[["means_v"]]) # Imprime la tabla contenida en el objeto
  "means_v."
  rk.header("Valores de la V de Aiken con intervalo de confianza (CI) para cada ítem",
  level = 2) # Presenta el subtítulo para el CI calculado para cada ítem.
  rk.results(aiken_vals[["v_ci"]], print.rownames = TRUE) # Imprime la tabla contenida en
  el marco de datos "v_ci".
})
```

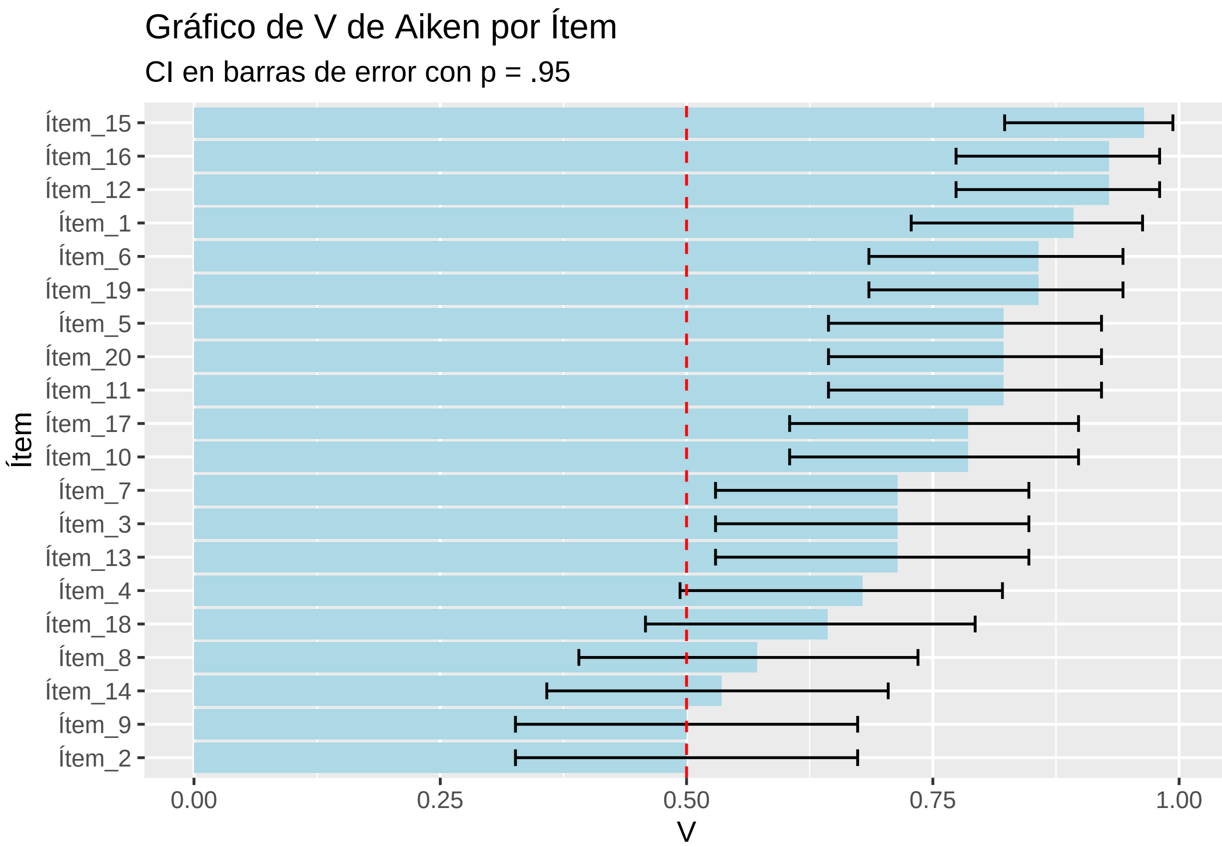

Figura 1: Gráfico de barras para la media del V de Aiken por cada Ítem y barras de error para su IC con línea de exclusión en el nivel 0.5

Por otro lado, el gráfico se puede visualizar en la ventada de salida de RKWard con;

```
rk.header(
  title = "Gráfico de barras para la media del V de Aiken por cada Ítem y barras de error
  para su IC con línea de exclusión en el nivel 0.5",
  parameters = aiken_vals[["parameters"]]
)
rk.graph.on(
  device.type = "PNG",
  width = 1024,
  height = 720,
  pointsize = 10.0,
  res = 150,
  bg = "transparent"
)
try({
  print(g)
})
rk.graph.off()
```

Primeramente se ingresa el título en HTML con la función `rk.header()`, se inicia el dispositivo gráfico con `rk.graph.on()`, con los argumentos de tipo "PNG", aunque también se encuentra diponible "JPG", el ancho de la imagen en `2024` por alto en `720` pixeles; Un tamaño de punto de `10`, con una resolución de `150` pixeles por pulga cuadrada (ppi por sus siglas en inglés) y fondo transparente. Para imprimir el resultado se usa el envoltorio `try({})`, para aplicar la función `print()` y finalmente se cierra el dispositivo gráfico con `rk.graph.off ()`.

4 Interpretación

Un valor de 0.75 se asocia con una calificación promedio de 4 en una escala de 5 puntos con opciones de respuesta que van del 1 al 5, o un buen ajuste al constructo deseado. En este caso, la determinación de los ítems para los cuales se acepta la hipótesis nula de $V_p = 0.75$ utilizando una tasa de error de tipo I de 0,05 se puede realizar determinando los ítems para los cuales el intervalo de confianza del 90% contiene 0,75 (todos los ítems excepto 1, 12, 15 y 16). Aunque el valor de criterio de 0,75 puede resultar demasiado estricto en la práctica, lo presentamos aquí con fines estrictamente didácticos para ilustrar la flexibilidad del intervalo de confianza de la puntuación en la prueba de hipótesis propuesta por Aiken (1980, 1985). Los investigadores que se encuentran en las etapas iniciales del desarrollo de la escala pueden optar por un valor de criterio más flexible (p. ej., $V_0 = 0.4$) o utilizar una tasa de error de tipo I más alta (p. ej., $\alpha = 0.10$), especialmente si el número de evaluadores expertos es reducido.

Referencias Bibliográficas

Aiken, Lewis R. 1980. "Content Validity and Reliability of Single Items or Questionnaires". *Educational and Psychological Measurement* 40 (4): 955–59. https://doi.org/10.1177/001316448004000419.

———. 1985. "Three Coefficients for Analyzing the Reliability and Validity of Ratings". *Educational and Psychological Measurement* 45 (1): 131–42. https://doi.org/10.1177/0013164485451012.

Newcombe, Robert G. 1998. "Two-sided confidence intervals for the single proportion: comparison of seven methods". *STATISTICS IN MEDICINE*, núm. 17: 857–72. https://wiley.bibliotecabuap.elogim.com/doi/10.1002/(SICI)1097-0258(19980430)17:8%3C857::AID-SIM777%3E3.0.CO;2-E.

Oksaviona, Vebby, Nur Islami, y Muhammad Nasir. 2023. "Development of PBL-Based Sound Wave Interactive Multimedia Using Lumi for Class XI High School Students". *Jurnal Penelitian Pendidikan IPA* 9 (10): 8008–15. https://doi.org/10.29303/jppipa.v9i10.4426.

Penfield, Randall D., y Peter R. Giacobbi Jr. 2004. "Applying a Score Confidence Interval to Aiken's Item Content-Relevance Index". *Measurement in Physical Education and Exercise Science* 8 (4): 213–25.

https://doi.org/10.1207/s15327841mpee0804_3.
Wilson, Edwin B. 1927. "Probable Inference, the Law of Succession, and Statistical Inference". *Journal of the American Statistical Association* 22 (158): 209–12. https://www.jstor.org/stable/2276774.