EJERCICIO DE ANÁLISIS DE ÍTTEMS *Y DE FIABILIDAD*.

ANÁLISIS DE LA CALIDAD DE LOS ÍTEMS   
*Y DE LA FIABILIDAD DE LA PRUEBA DE LÉXICO*

Este Informe tiene por objeto determinar la calidad de los ítems de una prueba de léxico e informar sobre la *fiabilidad* misma con vistas a que puedan adoptarse medidas para, en su caso, mejorar dicha prueba.

El Informe se estructura en diferentes partes: la primera de ellas, recoge el planteamiento general del estudio, incluyendo el problema que se ha estudiado y la metodología que se ha seguido para recoger datos empíricos acerca del mismo, detallando aspectos tales como diseño metodológico, muestra estudiada o las técnicas de recogida de datos; en la segunda parte, procedimiento de análisis, se explican las técnicas estadísticas utilizadas en el análisis de los datos, indicando con qué finalidad se ha empleado cada una de ellas; en la tercera parte, se presentan los resultados del estudio; y, finalmente, en la última parte se incluyen las conclusiones del mismo.

*1.- PLANTEAMIENTO DEL ESTUDIO*

El problema abordado en este estudio responde a la cuestiones ¿cuál es la dificultad y discriminación de los ítems que componen la prueba de léxico? *¿Cuál es su fiabilidad?* La metodología seguida para responder a esta cuestión se enmarca dentro de los métodos correlacionales y se concreta en un diseño metodológico que supone la aplicación de una prueba denominada “Léxico”, compuesta por 29 ítems, a un grupo de 117 sujetos.

*2.- PROCEDIMIENTO DE ANÁLISIS:*

El análisis de la calidad de los ítems se ha realizado a partir de dos indicadores: dificultad y discriminación. En el primer caso, el índice de dificultad de cada ítem se ha calculado considerando la frecuencia de aciertos sobre el total de la muestra, es decir, sin excluir omisiones o ítems no alcanzados. En el segundo caso, el índice de discriminación se ha calculado con el coeficiente de correlación biserial puntual (aplicación del coeficiente de correlación de Pearson a una variable medida en escala de intervalo y otra dicotómica), a partir de la respuesta a un ítem y al total de puntuación en la prueba. *La fiabilidad se ha culculado a partir del coeficiente alfa de Cronbach.*

*3.- RESULTADOS*

La tabla 3.1 recoge los índices de dificultad y discriminación para cada uno de los 29 ítems de la prueba.

|  |  |  |
| --- | --- | --- |
| ÍTEMS | ÍNDICE DE DIFICULTAD | ÍNDICE DE DISCRIMINACIÓN |
| 1 | **0, 97** | 0,26 |
| 2 | **0,99** | ***0,19*** |
| 3 | 0,93 | 0,29 |
| 4 | 0,86 | 0,36 |
| 5 | 0,94 | 0,29 |
| 6 | 0,92 | 0,35 |
| 7 | 0,93 | 0,28 |
| 8 | **0,97** | 0,27 |
| 9 | 0,86 | 0,27 |
| 10 | 0,79 | 0,34 |
| 11 | 0,80 | 0,34 |
| 12 | 0,95 | ***0,19*** |
| 13 | 0,92 | 0,39 |
| 14 | 0,78 | 0,40 |
| 15\* | **0,96** | ***0,11*** |
| 16\* | **0,97** | ***0,10*** |
| 17 | **0,97** | 0,19 |
| 18 | 0,92 | 0,27 |
| 19 | 0,94 | 0,26 |
| 20 | ***0,37*** | 0,47 |
| 21 | ***0,54*** | **0,53** |
| 22 | ***0,27*** | **0,51** |
| 23 | ***0,67*** | **0,49** |
| 24 | 0,71 | **0,48** |
| 25 | ***0,53*** | **0,52** |
| 26 | 0,92 | 0,27 |
| 27 | 0,91 | 0,36 |
| 28 | 0,91 | 0,26 |
| 29\* | **0,98** | ***0,10*** |

Tabla 3.1. Índices de dificultad y discriminación en la prueba “Léxico”.

Los resultados recogidos en la Tabla 3.1 muestran que la prueba “Léxico” tiene índices de dificultad y de discriminación heterogéneos. En el primer caso, con valores que oscilan entre 0,99 y 0,27 (los responden el 99% y el 27%, respectivamente) y el segundo, con valores entre 0,10 y 0,53.

Los índices de dificultad obtenidos llevan a cuestionar la calidad de los ítems 2 y 29, así como a revisar las razones por las que otros ítems, tales como el 1, 8, 15, 16 y 17, alcanzan valores superiores a 0,96. De igual modo, los índices de discriminación correspondientes a los ítems, 15, 16 y 29 están en torno a 0,10. Si elevamos ese valor al cuadrado (coeficiente de determinación) podemos comprobar que la varianza ítem total sólo queda explicada en un 1%, lo que implicaría una escasa capacidad de estos ítems para diferenciar el nivel léxico de los sujetos. Otros ítems que, por idéntica razón, deberían ser revisados son el 2, 12 y 17.

*La fiabilidad de la prueba de léxico obtenida a partir del coeficiente alfa de Cronbach es 0,71, lo que indica la mediocridad de la misma (los valores en pruebas de medición del léxico tienen coeficientes de fiabilidad en torno a 0,90) para diferenciar distintos niveles en la capacidad objeto de estudio. En este sentido, puede cuestionarse la consistencia interna de las mediciones realizadas y destacarse la existencia importes errores aleatorios de medida.*

*La tabla 3.2 muestra algunos datos de interés relativos a los ítems de la prueba de léxico. A partir de la correlación elemento-total corregida[[1]](#footnote-1) (valor corregido del índice de discriminación) se ha obtenido el valor del coeficiente alfa si se elimina el ítem. De ese modo, si eliminamos el ítem 15 y el ítem 16, aumentaría el valor de dicho coeficiente (0,713 y 0,712, respectivamente); al contrario, el valor del coeficiente alfa se vería sensiblemente reducido si eliminásemos los ítems 20, 21, 22, 23, 24 y 25, especialmente, los ítems 21 y 22 (alfa=0,689), lo que indica el alto valor de los mismos para diferenciar distintos niveles de léxico entre los sujetos que se han sometido a la prueba.*

|  | | | | |
| --- | --- | --- | --- | --- |
|  | Media de la escala si se elimina el elemento | Varianza de la escala si se elimina el elemento | Correlación elemento-total corregida | Alfa de Cronbach si se elimina el elemento |
| i\_1 | 23,23 | 9,432 | ,200 | ,707 |
| i\_2 | 23,21 | 9,592 | *,157* | ,710 |
| i\_3 | 23,27 | 9,293 | ,215 | ,705 |
| i\_4 | 23,34 | 9,028 | ,251 | ,703 |
| i\_5 | 23,26 | 9,324 | ,212 | ,706 |
| i\_6 | 23,28 | 9,162 | ,263 | ,702 |
| i\_7 | 23,27 | 9,311 | ,203 | ,706 |
| i\_8 | 23,22 | 9,445 | ,225 | ,706 |
| i\_9 | 23,34 | 9,234 | ,*160* | ,710 |
| i\_10 | 23,41 | 8,993 | ,213 | ,707 |
| i\_11 | 23,41 | 9,008 | ,212 | ,707 |
| i\_12 | 23,25 | 9,476 | ,*121* | ,710 |
| i\_13 | 23,28 | 9,092 | ,305 | ,700 |
| i\_14 | 23,42 | 8,820 | ,280 | ,701 |
| i\_15 | 23,24 | 9,593 | *,044* | **,713** |
| i\_16 | 23,22 | 9,619 | *,047* | **,712** |
| i\_17 | 23,22 | 9,532 | *,136* | ,709 |
| i\_18 | 23,27 | 9,328 | *,192* | ,707 |
| i\_19 | 23,26 | 9,359 | *,188* | ,707 |
| i\_20 | 23,83 | 8,509 | ,334 | ***,696*** |
| i\_21 | 23,66 | 8,297 | ,397 | ***,689*** |
| i\_22 | 23,93 | 8,465 | ,398 | ***,689*** |
| i\_23 | 23,53 | 8,478 | ,361 | ***,693*** |
| i\_24 | 23,48 | 8,548 | ,354 | ***,694*** |
| i\_25 | 23,66 | 8,329 | ,384 | ***,690*** |
| i\_26 | 23,28 | 9,315 | ,*185* | ,707 |
| i\_27 | 23,28 | 9,144 | ,274 | ,702 |
| i\_28 | 23,28 | 9,318 | *,170* | ,708 |
| i\_29 | 23,21 | 9,644 | *,067* | ,711 |

*Tabla 3.2 Correlaciones ítem-total y fiabilidad de la prueba (si se elimina el ítem).*

*4.-CONCLUSIONES*

La prueba léxico presenta una cierta variabilidad en la calidad de sus ítems, reflejada en los dos indicadores estudiados. En cuanto a su dificultad, predominan los ítems con índices que superan 0,80 y hay ítems en los que acierta la práctica totalidad de los alumnos, siendo muy pocos aquellos que responden correctamente menos del 50% de los alumnos. En cuanto a la discriminación, ésta se sitúa mayoritariamente en valores en torno a 0,30, aunque algunos ítems alcanzan sólo 0.10 puntos. Si consideramos que el valor de dicho índice oscila entre 0 y 1, la capacidad de los ítems de la prueba para diferenciar distintos niveles de léxico, entre los sujetos implicados, es muy limitada. *Este hecho se ve confirmado por el valor 0,71 del coeficiente de fiabilidad Alfa de Cronbach, el cual puede considerarse bastante mediocre.*

*Teniendo en cuenta el valor del índice de discriminación,* sería conveniente eliminar ítems como el 15, 16 y 29, dado que con esta medida mejoraría la calidad de la prueba Léxico, pues se aumentaría su fiabilidad[[2]](#footnote-2); *un análisis de la fiabilidad de la prueba ha venido a confirmar esta idea inicial, especialmente en el caso de los ítems 15 y 16*. De igual modo, sería oportuno revisar otros ítems, como el 2, 12 y 17, para tratar de comprender las razones que explican los bajos valores de dificultad y discriminación que han obtenido; *el análisis de la fiabilidad también ha venido a confirmar esta idea inicial, dado que los valores de correlación ítem-total de estos ítems (que no alcanzan el valor 0,20) están entre los más bajos de la Tabla 3.2*.

1. Al calcular la correlación item-total, en la puntuación total se ha descontado la puntuación de cada ítem. [↑](#footnote-ref-1)
2. Este concepto se explicará a lo largo del tema 4, cuando veamos cómo se calcula e interpreta el alfa de Cronbach. [↑](#footnote-ref-2)