

Recibido: 10-04-2020 / Revisado: 10-05-2020 / Aceptado: 15-06-2020 / Publicado: 03-07-2020

DOI: <https://doi.org/10.33262/exploradordigital.v4i3.1342>



## **Análisis estadístico de los resultados de aprendizaje específico Universidad Nacional de Chimborazo**

### *Statistical Analysis of Specific Learning Results National University of Chimborazo*

Héctor Salomón Mullo Guaminga.<sup>1</sup>, Jessica Alexandra Marcatoma Tixi.<sup>2</sup>, Natalia Alexandra Pérez Londo.<sup>3</sup> & Sonia Rodas Espinoza.<sup>4</sup>

#### **Abstract.**

The article deals with a topic of particular importance for university educational evaluation, focused on studying the quality of the Specific Learning Outcomes test using psychometric theory. With this intention, the work followed a non-experimental quantitative approach, with a descriptive scope, using the Classical Test Theory and the Item Response Theory. The results indicate that the tests have good quality in the design, construction, application, qualification and reporting phases; however, the verification phase is of poor quality, since piloting, calibration, and reagent selection are not performed. On the other hand, within the study of the items, quality problems are evident in difficulty and discrimination, while in reliability they showed consistency and precision in the scores obtained by the exam. Furthermore, the tests, as a whole, have a low power of discrimination between the different skill levels of the supporters. Finally, the results suggest that either of the two psychometric theories can be used to study the Specific Learning Outcome tests.

<sup>1</sup> Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Ecuador, [hmullo@esepoch.edu.ec](mailto:hmullo@esepoch.edu.ec);  
Instituto Nacional de Evaluación Educativa, Ecuador, [hector.mullo@evaluacion.gob.ec](mailto:hector.mullo@evaluacion.gob.ec)

<sup>2</sup> Universidad Nacional de Chimborazo, Facultad de Ingeniería, Ecuador, [jessica.marcatoma@unach.edu.ec](mailto:jessica.marcatoma@unach.edu.ec)

<sup>3</sup> Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Ecuador, [nperez@esepoch.edu.ec](mailto:nperez@esepoch.edu.ec)

<sup>4</sup> Universidad Nacional de Chimborazo, Facultad de Ingeniería, Ecuador, [srodas@unach.edu.ec](mailto:srodas@unach.edu.ec)

**Keywords:** quality, evaluation, university education, psychometry.

### **Resumen.**

El artículo trata sobre un tópico de particular importancia para la evaluación educativa universitaria, enfocado en estudiar la calidad de la prueba de Resultados de Aprendizaje Específico mediante la teoría psicométrica. Con esta intención, el trabajo siguió un enfoque cuantitativo no experimental, con un alcance descriptivo, utilizando la Teoría Clásica del Test y la Teoría de Respuesta al Ítem. Los resultados indican que las pruebas tienen buena calidad en las fases de diseño, construcción, aplicación, calificación y reportes; sin embargo, la fase de verificación presenta mala calidad, ya que no se realiza el pilotaje, la calibración y la selección de reactivos. Por otro lado, dentro del estudio de los reactivos, se evidencian problemas de calidad en dificultad y discriminación, mientras que en confiabilidad mostraron consistencia y precisión en las puntuaciones arrojadas por el examen. Además, las pruebas, como un todo, tienen bajo poder de discriminación entre los diferentes niveles de habilidad de los sustentantes. Finalmente, los resultados sugieren que se puede ocupar cualquiera de las dos teorías psicométricas para estudiar las pruebas de Resultado de Aprendizaje Específico.

**Palabras claves:** calidad, evaluación, educación superior, psicometría.

### **Introducción.**

La evaluación educativa hace referencia a la “acción permanente por medio de la cual se busca apreciar, estimar y emitir juicios sobre procesos de desarrollo del alumno o sobre los procesos pedagógicos o administrativos, así como sobre sus resultados, con el fin de elevar y mantener la calidad de estos” (Ministerio de Educación Nacional, 1997, p. 17). En correspondencia con esto, el Concejo de Aseguramiento de la Calidad de la Educación Superior (CACES) del Ecuador establece que las Instituciones de Educación Superior, tanto públicas como particulares, sus carreras y programas, deben someterse en forma obligatoria a la evaluación externa y a la acreditación (CACES, 2019a).

En particular, el proceso de evaluación obligatoria de carreras determinado en la normativa nacional del CACES se compone de dos etapas: a) la evaluación del Entorno del Aprendizaje y b) la evaluación de los Resultados de Aprendizaje. La evaluación del Entorno de Aprendizaje es un proceso que valora las condiciones académicas y físicas en las que se desarrolla la carrera, por otro lado, los Resultados de Aprendizaje evalúan a los estudiantes que se encuentran en el último año de la carrera en proceso de evaluación. El resultado final de los dos procesos determina la acreditación de una carrera, ubicando a la carrera en una de las tres categorías: acreditada, en proceso de acreditación y no acreditadas, de acuerdo con el artículo 46 del Reglamento de Evaluación, Acreditación y Categorización de las Instituciones

de Educación Superior (Codificado). Las carreras acreditadas certifican el cumplimiento de los criterios mínimos de calidad, necesarios para funcionar y garantizar una adecuada formación de sus estudiantes. En el caso de que una carrera no sea acreditada, el CACES determinará su suspensión, impidiendo a la institución la posibilidad de abrir nuevas promociones o cohortes, durante un periodo de diez años, a partir de la notificación respectiva (CACES, 2019b) (Cisneros, 2019).

La Universidad Nacional de Chimborazo (UNACH), en respuesta a las exigencias del CACES, se viene preparando para la evaluación de las carreras tanto en el Entorno del Aprendizaje como en los Resultados de Aprendizaje. Con respecto a esto último, la UNACH ha desarrollado e implementado las pruebas de Resultados de Aprendizaje Específico (RAE) en los estudiantes legalmente matriculados en el último año de las carreras, con aras en la mejor continua de la acreditación.

La construcción del instrumento de evaluación de la prueba RAE inicia con la capacitación a los docentes en la construcción de reactivos y la consideración de las asignaturas del campo praxis-profesional como la estructura de prueba. Posteriormente, al momento de la evaluación, se asignan (de forma aleatoria) 100 preguntas a cada sustentante, las mismas que luego se califican y, finalmente, se presentan los resultados de la evaluación en términos de porcentajes de aciertos (Cisneros, 2019). Por lo tanto, la prueba RAE cumple con la mayoría de los requerimientos para calificar como un buen proceso de evaluación educativa.

Al respecto Chávez y Saade (2009) indican que un buen proceso de evaluación educativa debe abarcar las fases de: diseño, construcción, verificación, ensamble, aplicación, calificación, reportes y mantenimiento. En este contexto, los autores Lord (1977), Chávez y Saade (2009), Baker y Kim (2017), Lord (2012) y CENEVAL (2012) indican que la fase de verificación y, como parte de esta, la revisión cuantitativa o análisis estadístico de los reactivos es transcendental, debido a que es una forma de evidenciar que la elaboración del instrumento de evaluación va por buen camino (control de calidad) y abre una especie de diálogo entre quienes diseñan la prueba y elaboran los reactivos y quienes estudian la evidencia empírica que resulta de aplicar la prueba en situaciones reales.

La prueba RAE evidencia dos deficiencias en la fase de verificación cuantitativa. En primer lugar, los reactivos no son piloteados y, por tanto, no existe una calibración en términos de dificultad, discriminación y confiabilidad. En segundo lugar, no se sigue una metodología para la selección de reactivos de un banco de ítems para la construcción de la prueba. Esto es preocupante debido a que, sin un adecuado análisis estadístico de los reactivos y posterior selección de estos para la prueba, no se puede garantizar que el instrumento de evaluación mantenga consistencia en sus parámetros psicométricos a lo largo del tiempo y que se

conserve una métrica supervisada, ya sea mediante la Teoría Clásica del Test (TCT) o la Teoría de Respuesta al Ítem (TRI).

Respondiendo a la importancia de la fase de verificación cuantitativa y sus deficiencias en la prueba RAE, esta investigación se centra en el estudio de la calidad del instrumento de evaluación, específicamente de la Carrera de Ingeniería Agroindustrial, mediante TCT y TRI. Además, se espera generar, en posteriores aplicaciones de la RAE, un mayor rigor técnico y metodológico en la fase de verificación y, por tanto, en la construcción de los instrumentos, empleados de manera que, por un lado, permitan tomar decisiones relacionadas directamente con la calibración y selección de reactivos de la prueba, y, por otro, ayuden a reconocer la calidad del instrumento completo.

**Teoría Clásica del Test:** La TCT, siguiendo a Meneses et al. (2013) junto con Chávez y Saade (2009), se concentra en el análisis del examen como una unidad y se basa en el modelo lineal clásico propuesto por Spearman, que articula el proceso de medida, definiendo tres conceptos fundamentales: la puntuación empírica ( $X$ ), la puntuación verdadera ( $V$ ) y el error de medida ( $E$ ), según la ecuación:

$$X = V + E$$

Algunos indicadores de TCT son la dificultad, la discriminación y la fiabilidad de los reactivos.

**Teoría de la respuesta al ítem:** La TRI plantea la existencia de una relación entre las puntuaciones en la variable latente (habilidad de los sustentantes) y la probabilidad de acertar cada ítem, introduciendo tres funciones matemáticas para modelar adecuadamente esta relación, las cuales se conocen como Curvas Características del Ítem (CCI). A continuación, se presenta el modelo logístico de dos parámetros<sup>5</sup>:

$$P_i(\theta) = \frac{1}{1+e^{-a_i(\theta-b_i)}} \quad i = 1, 2, \dots, k$$

donde  $P_i(\theta)$  es la probabilidad de que un examinado con habilidad  $\theta$  responda correctamente el ítem  $i$ ,  $b_i$  es la dificultad,  $a_i$  la discriminación y  $e$  la constante neperiana. Además, tanto  $\theta$ ,  $a$  y  $b$  toman valores entre  $-\infty$  y  $+\infty$ .

Algunas transformaciones importantes de la escala de habilidad es la Curva Característica de la Prueba (CCP), que estima el número de respuestas correctas esperado para un sustentante con habilidad  $\theta$ , la Función de Información del Ítem (FII) o de la prueba (FIP) permite, entre

---

<sup>5</sup> Siguiendo a Meneses et al. (2013) junto con Chávez y Saade (2009)

otras cosas, seleccionar reactivos para el ensamble de versiones, comparar pruebas y definir la precisión en la estimación de una habilidad específica.

La calidad de la estimación con los modelos logísticos se mide con un indicador de ajuste del modelo a los datos. En este sentido, se tienen dos indicadores de ajuste: uno interno (infit) y otro externo (outfit). El primero mide la sensibilidad al comportamiento inesperado que afecta a los ítems que presentan un nivel de dificultad lejano del nivel de habilidad de una persona y el segundo mide la sensibilidad en relación con la cercanía del nivel de habilidad de la persona (Tristan, 1998). Los valores de ajuste óptimos de los ítems deben encontrarse entre 0.5 y 1.5 (Linacre, 2003).

### **Metodología.**

**Participantes:** Los participantes de las pruebas RAE fueron 55 y 63 sustentantes en los periodos académicos 2018-2019 y 2019-2019 respectivamente, quienes estaban legalmente matriculados en la Carrera de Ingeniería Agroindustrial de la UNACH y cursaban el último año de carrera.

Hubo, en promedio, 63 y 78 aciertos en los ítems de los periodos 2018-2019 y 2019-2019, respectivamente. En estos casos se evaluaron 100 preguntas de selección múltiple con tres a cinco opciones de respuesta, de un banco de 1 705 ítems para 2018-2019 y 3 338 en 2019-2019.

**Proceso:** La recolección de los datos se consiguió a partir de una solicitud al Director de Evaluación y Aseguramiento de la Calidad Institucional de la UNACH. Una vez recolectados los datos se hizo una selección de ítems calibrables, ya que, por ejemplo, para el caso del periodo 2018-2019 se aplicaron 100 preguntas a cada sustentante, de un banco de 1 705 ítems. Esto produjo que hubiera reactivos que no fueron utilizados y otros que solo se utilizaron unas pocas veces, generando así un problema al momento de calibrar los ítems, por lo tanto, se escogieron aquellos ítems que fueron respondidos por al menos ocho sustentantes, cuidando que no todos los examinados hubieran respondido correctamente o incorrectamente el ítem. Después de aplicar este proceso quedaron 154 ítems calibrables, donde se imputaron los valores faltantes utilizando un remuestreo de tamaño 1 000, de una distribución Bernoulli con probabilidad de éxito igual a la proporción de aciertos. Para el periodo 2019-2019, en el que se aplicaron 3 338 ítems, se ocupó el mismo procedimiento del periodo anterior, quedando así 84 preguntas calibrables. Luego se efectuó el estudio estadístico pertinente mediante las teorías psicométricas TCT y TRI.

En este marco, el trabajo siguió un enfoque cuantitativo, con un alcance descriptivo, ya que se estudió la calidad de las pruebas e ítems mediante TCT y TRI. El diseño de investigación fue no experimental de corte longitudinal, debido a que se observaron los resultados de las pruebas en el ambiente natural de los sustentantes en dos periodos académicos.

Para el estudio estadístico mediante TCT y TRI se utilizaron los paquetes “irtoys” y “TAM”, de los autores Partchev, Maris y Hattori (2017) y Robitzsch, Kiefer y Wu (2019) del software estadístico R, con el que se obtuvieron las estimaciones de las dificultades, discriminaciones y alfa de Cronbach para TCT y las funciones de información y curvas carecterísticas mediante el modelo logístico de dos parámetros en el caso de TRI.

Después de la calibración, los ítems se catalogaron como “aceptados” según dificultad, districriminación y dificultad-discriminación cuando los valores estaban dentro de los intrevalos que muestra la tabla 1. La confiabilidad se obtuvo a partir del coeficiente alfa de Cronbach en TCT y para el ajuste al modelo de dos parametris en TRI se ocuparon los indicadores infit y outfit.

**Tabla 1.** Dictamen TCT y TRI de aceptación de ítems

TCT		TRI	
Dificultad	Discriminación	Dificultad	Discriminación
(0.3, 0.7)	(0.1, 1)	(-3, +3)	(0.1, 2.8)

**Fuente:** Elaboración propia.

## Resultados.

**Estudio de ítems TCT:** Se encontró que para el periodo 2018-2019, 44.81% (69), 35.06% (54) y 17.53% (27) de los ítems fueron aceptados según dificultad, discriminación y dificultad-discriminación. Para el periodo 2019-2019 la situación mejoró y se tuvo que 80.95% (68), 44.05% (37) y 38.10% (32) fueron aceptados según dificultad, discriminación y dificultad-discriminación. En general, las dos pruebas mostraron bajos porcentajes de ítems que cumplen con los criterios.

La distribución del índice de dificultad de los ítems mostró que para el periodo 2018-2019, el 36% (56) de ítems son altamente dificultosos (valores menores a 0.40), mientras que el 56.49% (87) tuvieron dificultades bajas (valores mayores a 0.60). Tan solo el 7.51% tiene dificultades moderadas (puntajes entre 0.40 y 0.60). Esto se aprecia en la Figura 1 y además, evidencia la falta de ítems en algunos rangos de dificultad, por lo que la prueba es una dicotomía entre ítems fáciles y otros difíciles, lo que no permite una evaluación adecuada.

Por otro lado, para el periodo 2019-2019, en la Figura 1 se observa una mejor distribución de los ítems en el rango de dificultad, aunque no hubo ítems con dificultades menores a 0.30 y el 85.00% (71) de los ítems tuvieron dificultades bajas.

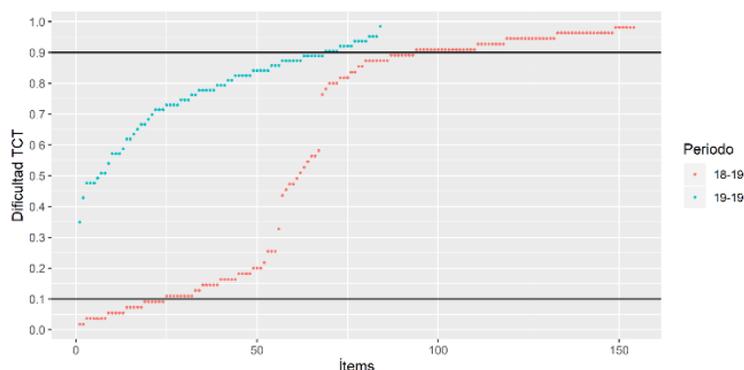
Con relación a la discriminación de los ítems (ver Figura 2) para el periodo 2018-2019, el 64.94% (100) tuvieron discriminaciones inferiores a 0.20 y casi todos los ítems restantes no superaron una discriminación de 0.42, a excepción de uno que tiene un valor de 0.54. Esto

quiere decir que, en general, las discriminaciones de los ítems son relativamente bajas. Algo similar ocurre en el periodo 2019-2019, donde el 55.95% (43) tuvo discriminaciones inferiores a 0.20 y ningún ítem superó el valor de 0.54 en discriminación.

Sobre calidad las medidas del coeficiente alfa de Cronbach para las pruebas RAE de los periodos 2018-2019 y 2019-2019 fueron muy similares (0.65 y 0.62, respectivamente) e indicaron que las dos pruebas son altamente confiables; es decir, hubo consistencia y precisión en los valores arrojados por el examen.

En lo referente a validez de contenido, se encontró que en los dos procesos de evaluación RAE se siguieron los caminos que la psicometría enseña para hacer pruebas educativas y asegurar su validez de contenido.

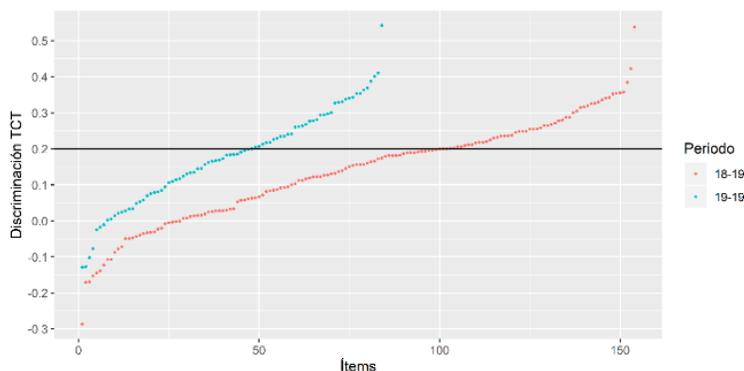
**Figura 1.** Dificultad pruebas RAE por ítem, periodos 2018-2019 y 2019-2019



**Fuente:** Elaboración propia.

**Nota:** Las líneas horizontales determinan los dictámenes.

**Figura 2.** Discriminación pruebas RAE por ítem, periodos 2018-2019 y 2019-2019

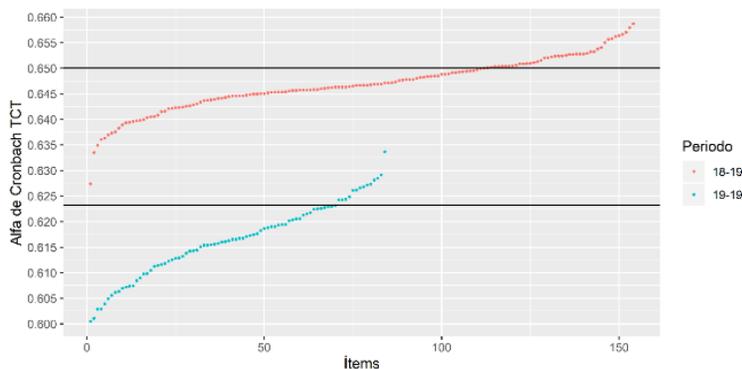


**Fuente:** Elaboración propia.

**Nota:** La línea horizontal determina el dictamen.

**Estudio de ítems TRI:** Para el periodo 2018-2019 el 33.12% (51), 38.31% (59) y 18.18% (28) de los ítems fueron aceptados según su dificultad, discriminación y dificultad-discriminación. Mientras que para el periodo 2019-2019, 48.81% (41), 44.05% (37) y 28.57% (24) de los reactivos fueron aceptados según su dificultad, discriminación y dificultad-discriminación. Este panorama es similar al mostrado según TCT.

**Figura 3.** Alfa de Cronbach pruebas RAE por ítem, periodos 2018-2019 y 2019-2019



**Fuente:** Elaboración propia.

**Nota:** Las líneas horizontales indican los coeficientes alfa de Cronbach de las pruebas.

De aquí en adelante, para mejorar la presentación de los resultados se estudiaron los ítems aceptados según dificultad-discriminación. De estos, las CCI (ver Figura 4) para el periodo de evaluación 2018-2019 muestran que algunas curvas de los ítems se centraron en niveles altos de habilidad ( $\theta > 1$ ) con discriminaciones también altas ( $\bar{a} = 1.22$ ), mientras que otros ítems están en niveles de habilidad bajo ( $\theta < -1$ ), con discriminaciones altas ( $\bar{a} = 1.26$ ). Por otro lado, unos pocos ítems se centran en un nivel de habilidad moderado ( $-1 \leq \theta \leq 1$ ) con dicriminaciones bajas ( $\bar{a} = 0.54$ ).

Para el periodo 2019-2019 hubo curvas que se centraron en niveles de habilidad bajo con discriminaciones moderadas ( $\bar{a} = 0.81$ ), y otras curvas se ubicaron en habilidades moderadas con discriminaciones en el mismo sentido ( $\bar{a} = 0.80$ ). No se encontraron curvas centradas en habilidades altas. Por lo tanto, la prueba 2018-2019 tuvo mayoritariamente ítems fáciles y difíciles, con buena capacidad para discriminar entre sustentantes que manifestaron mayor y menor conocimiento del contenido evaluado; mientras que para el periodo 2019-2019 los ítems fueron fáciles y moderados, con discriminaciones discretas.

En cuanto a la FII, para el periodo 2018-2019 (ver Figura 5) se evidencia que algunos ítems dan buena información para los niveles de habilidad bajo y alto. Además, se observó que la mayor información (1.74) se da para el nivel de habilidad -2.4; por otro lado, la evaluación

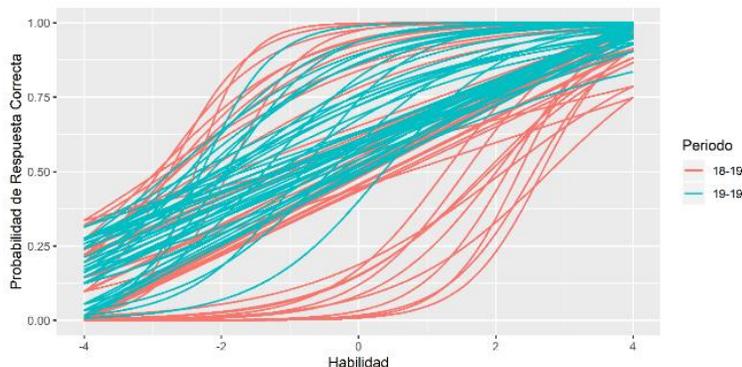
del periodo 2019-2019 indicó que los ítems dieron extensa información en los niveles de habilidad bajo y moderado, y la mayor información (0.97) se observó en el nivel de habilidad -2.32.

Por otro lado, se estudiaron las pruebas RAE como un todo, sin excluir aquellos ítems que no cumplieron con los dictámenes establecidos. Esto fue evidente al observar en las FIP's de los periodos 2018-2019 y 2019-2019 (ver Figura 6) que estas pruebas evaluaron con más precisión a aquellos estudiantes con niveles bajos y altos de habilidad. Este comportamiento no obedece a ningún tipo de prueba estandarizada, por ejemplo, la prueba de detección, máxima y de amplio rango.

En el mismo sentido, las CCP del periodo 2018-2019 (ver Figura 7) en general muestran una pendiente pequeña, lo cual no permite una adecuada discriminación entre aquellos sustentantes con mayor y menor habilidad; por ejemplo, con habilidades de -2 y 2 se tuvo una esperanza de 82 y 103 aciertos, respectivamente; es decir, solo hubo 21 aciertos de diferencias entre 154 preguntas. De la misma manera, para el periodo 2019-2019 se observó que, con habilidades de -2 y 2, hubo una esperanza de 56 y 68 aciertos, con una diferencia de 12 aciertos de un total de 84 reactivos. En consecuencia, esta prueba también tuvo un bajo poder de discriminación entre los diferentes niveles de habilidad.

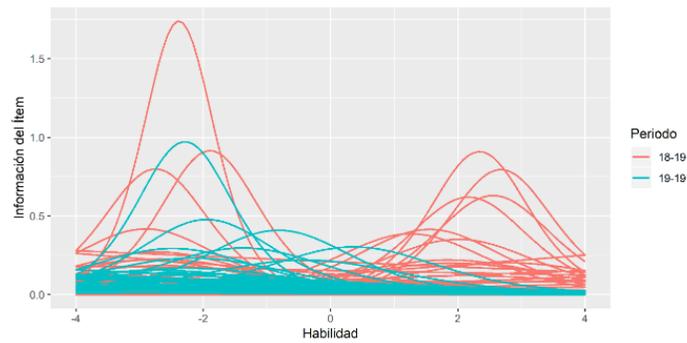
Siguiendo el orden de ideas, la calidad de la estimación TRI, la cual indica con qué precisión o previsibilidad se ajustan los datos al modelo, mostró que para el periodo 2018-2019 las medidas de ajuste de los reactivos tuvieron una media de 1.00 (d.e=0.12) y 0.90 (d.e=0.23) para infit y outfit, respectivamente. En el periodo 2019-2019 los valores tuvieron un promedio de 0.99 (d.e=0.09) en infit y 0.97 (d.e=0.15) para outfit. Esto indicó que hubo una buena calidad de la estimación del modelo, es decir, el modelo de dos parámetros fue una aproximación adecuada para el estudio de estas pruebas.

**Figura 4.** CCI o Probabilidad de Respuesta Correcta del Ítem según habilidad, periodos 2018-2019 y 2019-2019



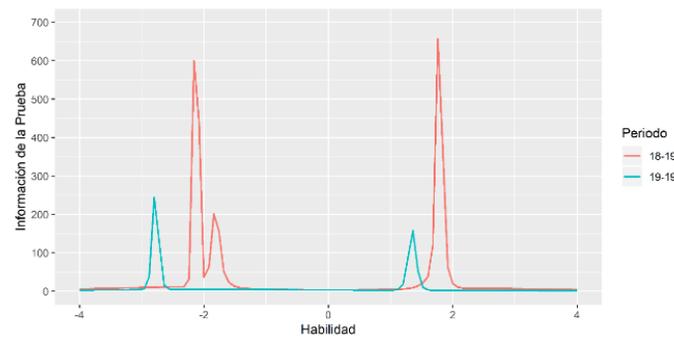
**Fuente:** Elaboración propia.

**Figura 5.** FII, Función de Información del Ítem por habilidad, periodos 2018-2019 y 2019-2019



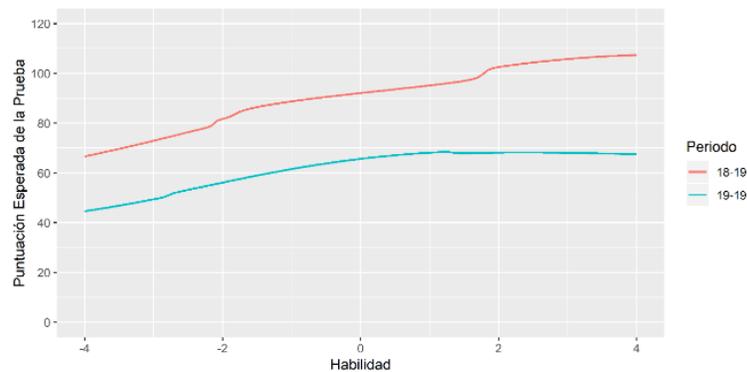
**Fuente:** Elaboración propia.

**Figura 6.** FIP, Función de Información de la Prueba según habilidad, periodos 2018-2019 y 2019-2019



**Fuente:** Elaboración propia.

**Figura 7.** CCP, Puntuación Esperada de la Prueba según habilidad, periodos 2018-2019 y 2019-2019



**Fuente:** Elaboración propia.

## Conclusiones:

- En virtud de los resultados alcanzados, se concluye que, en general, la calidad de los instrumentos utilizados en las pruebas RAE de los periodos 2018-2019 y 2019-2019 es buena en las fases de diseño, construcción, aplicación, calificación y reportes; sin embargo, en la fase de verificación, la calidad es mala, ya que no se realiza el pilotaje, calibración y selección de reactivos de las pruebas. Además, la fase de ensamble no presenta buena calidad, debido a que utiliza información de la fase de verificación e integra la versión operante de la prueba, tomando como insumo los mejores reactivos (seleccionados en la fase de verificación) en el mismo sentido. El no tener buena calidad en la fase de ensamble imposibilita una posterior equiparación de los instrumentos en el tiempo (Downing y Haladayna, 2006). Adicional a esto, la fase de mantenimiento presenta mala calidad, debido a que después de comunicar los resultados no hay una etapa de actualización del contenido del instrumento, es decir que no se escogen los mejores ítems para agregarlos al banco de ítems de la prueba. Por lo tanto, las pruebas RAE necesitan ser revisadas, asumiendo su naturaleza experimental, mediante estudios que se enfoquen en estas fases. Esto sería un aporte significativo al conocimiento de las pruebas.
- Se deben tenerse en cuenta algunas limitaciones. Primero, aunque el proceso de imputación de valores en la matriz de respuestas fue adecuado y consistente con la información de las pruebas RAE, se deben generar menos ítems y aplicarlos a la mayoría o al total de sustentantes. En segundo lugar, debido a que este estudio se centró en los estudiantes del último año de la Carrera de Agroindustrial, no hay un número adecuado de sustentantes para una buena calibración de los reactivos, esto es de 200 a 500 sustentantes en el caso de TCT y más de 500 en TRI. En consecuencia, la indagación de la sensibilidad de las calibraciones con diferentes números de sustentantes en el contexto de las pruebas RAE, utilizando simulación estadística, es una línea importante para posteriores estudios.
- Finalmente, es importante indicar que, si las pruebas RAE no mejoran su calidad en las fases de verificación, ensamble y mantenimiento, seguramente seguirán persistiendo los problemas detectados en esta investigación.

## Referencias bibliográficas.

Baker, F., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R*.

Barbero, M., Vila, E., & Holgado, F. (2015). *Psicometría*. Madrid: Sanz y Torres.

CACES. (2019a). *Evaluación de Carreras*. Obtenido de <https://www.caces.gob.ec/web/ceaaces/carreras>

- CACES. (2019b). RESOLUCIÓN No. 013-SE-06-CACES-2019. Obtenido de <https://www.caces.gob.ec/documents/20143/488487/RESOLUCI%C3%93N+No.+013-SE-06-CACES-2019.pdf/469ed064-0d77-71f4-6bba-758dbdee64fc>
- CENEVAL. (2012). Manual Técnico enlace Media Superior 2008-2010.
- Chávez, C., & Saade, A. (2009). Procedimientos básicos para el análisis de reactivos Cuaderno técnico 8. México.
- Cisneros, S. (2019). Informe General De Evaluación De Resultados De Aprendizaje Específico (RAE). UNACH, Riobamba. Obtenido de [http://www.unach.edu.ec/wp-content/D.E.A/3\\_EVALUACIO%CC%81N%20DE%20RESULTADOS%20DE%20APRENDIZAJE%20ESPECI%CC%81FICOS/3\\_Informe%20RAE%20Institucional%2018\\_19.pdf](http://www.unach.edu.ec/wp-content/D.E.A/3_EVALUACIO%CC%81N%20DE%20RESULTADOS%20DE%20APRENDIZAJE%20ESPECI%CC%81FICOS/3_Informe%20RAE%20Institucional%2018_19.pdf)
- Downing, S., & Haladayna, T. (2006). Handbook of test development. Lawrence Erlbaum Associates.
- Linacre, M. (2003). A user's guide to Winsteps. Chicago: Mesa Press.
- Lord, F. (1977). Practical Applications of Item Characteristic Curve Theory. Journal of Educational Measurement, 14(2).
- Lord, F. (2012). Applications of Item Response Theory to Practical Testing Problems. Routledge.
- Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L., Turbany, J., & Valero, S. (2013). Psicometría.
- Ministerio de Educación Nacional. (1997, p. 17). Evaluación en el aula y más allá de ella. Bogotá.
- Partchev, I., Maris, G., & Hattori, T. (2017). A Collection of Functions Related to Item Response Theory (IRT).
- Robitzsch, A., Kiefer, T., & Wu, M. (2019). Test Analysis Modules.
- Rodríguez, O., Casas, P., & Medina, Y. (2005). Análisis Psicométrico De Los Exámenes De Evaluación De La Calidad De La Educación Superior (ECAES) En Colombia. Avances en Medición, 3, 153-172.
- Tristan, A. (1998). Análisis de Rasch para todos. (CENEVAL, Ed.) México.

**PARA CITAR EL ARTÍCULO INDEXADO.**

Mullo Guaminga, H. S., Marcatoma Tixi, J. A., Pérez Londo, N. A., & Rodas Espinoza, S. (2020). Análisis Estadístico de los Resultado de Aprendizaje Específico Universidad Nacional de Chimborazo. *Explorador Digital*, 4(3), 214-225.  
<https://doi.org/10.33262/exploradordigital.v4i3.1342>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Explorador Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Explorador Digital**.

