



EXAMEN GENERAL DE CONOCIMIENTOS PARA LA LICENCIATURA EN PSICOLOGÍA EDUCATIVA, PLAN 09, UNIVERSIDAD PEDAGÓGICA NACIONAL. INDICADORES PSICOMÉTRICOS

Cuauhtémoc Gerardo Pérez López
cgperez@upn.mx

Pedro Bollás García
pbollas@hotmail.com,

María Imelda González Mecalco
piowacwe@yahoo.com.mx

Alba Yanalte Álvarez Mejía
yanicuau@yahoo.com.mx,

Rocío Castro Galván
chio026@yahoo.com.mx

Resumen

El presente es un informe técnico de la primera aplicación del examen general de conocimientos de la Licenciatura en psicología educativa. Este examen se aplica para que los egresados de dicho programa educativo accedan al grado. Es una prueba de evaluación objetiva, la cual consta de 300 reactivos distribuidos en nueve áreas, que corresponden a las siete líneas curriculares del plan de estudios 09. Dicho instrumento tiene la particularidad de contar con únicamente tres opciones de respuesta, una es correcta y dos son distractoras. Participaron en la aplicación 76 egresados de la licenciatura, 48 de ellos fueron del nuevo plan de estudios y el resto son egresados que se formaron con el plan 90. Con el objetivo de iniciar el estudio y seguimiento de la funcionalidad de los reactivos, se calcularon dos de los indicadores psicométricos de este tipo de pruebas: dificultad y discriminación. Aun cuando son pocos los participantes, con los hallazgos encontrados se puede decir que el examen es un buen instrumento para identificar a los egresados con los conocimientos curriculares suficientes para acceder al grado. Asimismo, se observó que en la mayoría de los reactivos los indicadores psicométricos tienen valores adecuados. Sin embargo, una buena cantidad de ellos deben ser revisados, a la luz de la distribución de los distractores. Es necesario esperar los resultados de aplicaciones posteriores para poder decidir acerca de los contenidos evaluados en el examen. Sin embargo, se cuenta con información para informar a los profesores diseñadores de reactivos sobre los criterios a considerar para continuar trabajando en esta tarea.

Palabras clave. Titulación, examen de conocimientos, pruebas objetivas, indicadores psicométricos



EVALUACIÓN DEBATE 2014



Introducción

El plan de estudios de la licenciatura en psicología educativa fue evaluado por el CIEES correspondiente en 2006, como un programa de excelencia. Dentro de las recomendaciones derivadas de dicho proceso se propuso: “Revisar, y si fuera el caso adecuar el plan de estudios de la licenciatura en psicología educativa.” (CIEES, 2006, pág. 13). A partir de ello, en 2008 se inició un proyecto para rediseñar el plan de estudios, el cual se concluyó en 2009.

Una segunda recomendación está relacionada con el índice de titulación del programa educativo. Así, se sugiere: “Abrir la opción de titulación por examen general de conocimientos a todos los alumnos que han concluido los créditos del programa y que cubren con los requisitos de titulación” (pág. 19). La estrategia en el rediseño del plan de estudios fue considerar el EGC una opción regular para acceder al grado. De ese modo, la acción específica para atender esta situación fue modificar el instructivo de titulación (González, García, Pérez y Soriano, 2013). En este documento se menciona que los egresados tienen dos opciones para titularse: elaborar tesis o presentar el examen general de conocimientos.

En el instructivo se define el examen general de conocimientos de la siguiente manera: “Consiste en una prueba objetiva de opción múltiple, a través de la cual el egresado muestra su capacidad en el manejo de los contenidos curriculares del plan de estudios y su aplicación en diversas problemáticas educativas”; el cual tiene como objetivo formativo: “Demostrar ante una prueba objetiva habilidades de análisis, síntesis y aplicación de los conocimientos adquiridos durante su formación” (pág. 46).

Bajo esa condición la responsable del programa educativo nombró una subcomisión de la comisión de titulación, cuyo objetivo fue diseñar el examen y ofrecerlo a los egresados de la primera generación del plan de estudios 09. Este grupo de trabajo convocó a los integrantes del colegio de profesores a participar en esta tarea. En un trabajo posterior y más amplio se describirá a detalle el proceso para el diseño del examen. Este proceso se guió en la propuesta de doce pasos elaborada por Downing (2006) y la de Backhoff et al (2006). A grosso modo se mencionan las tareas en la que el colegio de profesores de la licenciatura se involucró para desarrollar las distintas fases esta tarea.

Con base en la definición y el objetivo formativo del EGC los integrantes de las siete líneas curriculares se dieron a la tarea de seleccionar los contenidos que se evaluaron en el examen. Estos fueron elegidos a partir de los objetivos de aprendizaje propuestos en los programas de los distintos cursos de la licenciatura. Con esta información se elaboró la tabla de especificaciones, fundamento para diseñar los reactivos. Se realizó un estudio piloto para cuantificar en primera instancia la claridad de las bases y opciones de los reactivos, así como determinar el tiempo que se requiere



para la aplicación. Con los documentos que en las líneas curriculares se elaboró la guía de estudio para los sustentantes.

Todo proceso de evaluación educativa involucra el diseño y la construcción, la aplicación y la calificación de los instrumentos para medir el aprendizaje, el proceso, las instituciones o la docencia. La principal función de la evaluación es encontrar información que permita una mejor toma de decisiones. En el caso de los instrumentos de alto impacto como los exámenes de admisión, certificación o titulación, con los que el resultado de la medición tiene un importante impacto social, la certeza de la decisión debe ser, en la medida de lo posible, garantizada.

Durante muchas décadas en México se aplicaron los exámenes de admisión en diversos niveles educativos sin realizar procedimiento alguno para cuantificar la validez de las decisiones. Con la aparición del CENEVAL esto ha cambiado de manera gradual. Esta institución así como algunas universidades del país se han preocupado por utilizar instrumentos mejor contruidos que cumplan con los estándares establecidos en diversas instancias internacionales (Backhoff et al, 2000).

De acuerdo con los autores, en ese año no existían organismos nacionales que se encargaran de establecer normas sobre los criterios para incrementar la calidad de los exámenes, en consecuencia, el desarrollo de de pruebas válidas y estandarizadas era nulo y, por tanto tampoco había publicaciones que informaran acerca de los indicadores psicométricos de las pocas pruebas que se aplicaban.

En la literatura especializada, se mencionan tres los indicadores psicométricos que se calculan para medir la calidad de la prueba que se aplica: dificultad-facilidad, discriminación y la distribución de respuestas entre las opciones falsas del reactivo. Para fines de este trabajo se describirá cómo se obtienen los dos primeros. Así, la dificultad-facilidad representa la proporción de examinados que responden correctamente el reactivo. De este modo si la proporción es mayor, la dificultad será menor. El indicador de discriminación es la medida que permite medir el nivel de funcionamiento de cada reactivo para diferenciar entre los examinados con mayores puntajes y quienes obtuvieron las calificaciones más bajas. El valor **D** que se obtiene en este indicador varía de -1 hasta 1. Los valores se interpretan con base en la siguiente tabla.

Tabla 1. Poder de discriminación de los reactivos según su valor D

Índice de discriminación	Calidad del reactivo	Recomendaciones
> 0.39	Excelente	Conservar
0.30 – 0.39	Buena	Posibilidades de mejorar
0.20 – 0.29	Regular	Necesidad de revisar
0.0 - 0.19	Pobre	Descartar o revisar a profundidad
> -0.01	Pésima	Descartar definitivamente



Tomado de Backhoff et al, (2000).

Para el diseño del EGC plan 09, el grupo de trabajo tomó en cuenta la experiencia con el diseño del examen para el plan 90 (Pérez, Hernández, Álvarez y Rubio, 2011), por lo cual se determinó que el del plan 09 contendría 300 reactivos. Para el diseño de los reactivos se consideraron las directrices que proponen Haldayana, Downing y Rodrigues (2002); Moreno, Martínez y Muñiz (2004) y CENEVAL (2000; 2011). En el proceso del diseño, se asumió una decisión de la mayor relevancia: elaborar reactivos con solo tres opciones de respuesta. Esto se decidió con base en los hallazgos de Baghaei y Amrahi (2011), Bruno y Dirkzwager (1995), Haladyna y Downing (1993) y Rodrigues (2005), quienes concluyen que para los exámenes de elección múltiple, tres opciones distractoras son suficientes.

Un elemento importante para haber decidido diseñar un examen de este tipo con tres opciones de respuesta es que en esos estudios se concluye que, con esta alternativa los indicadores psicométricos se alteran poco. Es decir, el examen tiene una facilidad mayor, pero el índice de discriminación no sufre modificaciones.

Asimismo, se menciona que la clave para el diseño de reactivos no es la cantidad de opciones, sino la calidad de los distractores. En los trabajos citados se argumenta y ejemplifica que en este tipo de exámenes elaborar la cuarta opción es, la mayor parte de las ocasiones, demasiado complicado. Esto porque se requiere elegir un concepto o elaborar una frase que sea del mismo campo semántico y sea una opción factible. Debido a lo anterior, mencionan los autores, esa frase o concepto resulta ser obviamente falsa, por lo que los sustentantes la eligen con una reducida frecuencia.

Por otro lado, reducir el número de opciones permite que en el examen se incluyan más reactivos, es decir evaluar mayor cantidad de contenidos, lo cual redundará en una mayor eficiencia para los sustentantes de altas calificaciones y reduce la eficiencia para quienes obtienen calificaciones bajas.

El presente es un informe técnico cuyo objetivo es describir las características psicométricas de la primera aplicación del EGC plan 09 de la Licenciatura en psicología educativa de la Universidad Pedagógica Nacional.

Método

El examen consta de 300 reactivos cada uno con tres opciones, solo una de ellas es la correcta. La distribución por línea curricular se observa en la tabla 1. Es importante mencionar que el total para cada área se asignó de acuerdo con la proporción de cursos-créditos que cada línea curricular contiene en los seis primeros semestres de la licenciatura. Asimismo, debe aclararse que son siete las líneas curriculares en las que se agrupan los distintos



cursos de la licenciatura. Sin embargo para organizar los reactivos en el examen, la línea Desarrollo y aprendizaje se dividió en dos secciones y, de la misma manera, la Metodológica se partió en la que se evaluó contenidos de Estadística y la segunda, contenidos de Metodología. De este modo, el examen se conformó por nueve áreas o bloques de reactivos agrupados de acuerdo con el contenido a evaluar.

Tabla 2. Distribución de reactivos por línea curricular

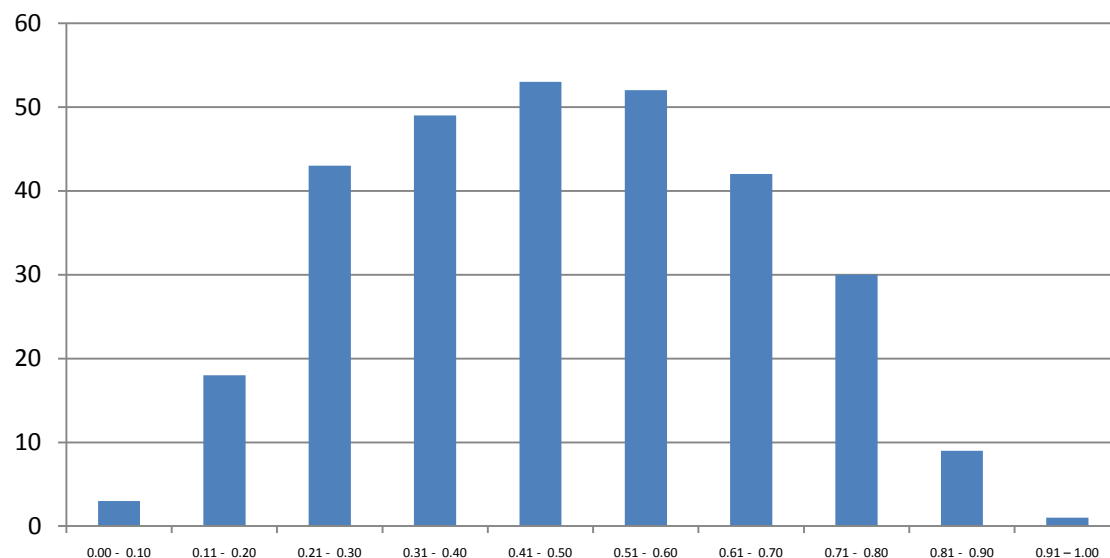
Línea curricular	Cantidad de reactivos
Desarrollo	45
Aprendizaje	20
Socioeducativa	20
Psicopedagógica	40
Educación inclusiva	30
Metodológica	35
Estadística	25
Diseño de materiales	40
Psicología social de la educación	45

En esta primera aplicación participaron 76 egresados de la licenciatura. De ellos 48 eran de la primera generación del plan 09 y el resto fueron formados en el plan anterior (plan 90). La distribución de los egresados por sexo es similar a la existente en la población estudiantil del programa educativo, el 87% es población femenina. Las edades mayores son las de los participantes de las generaciones con un periodo mayor de su egreso. Finalmente, el promedio de egreso es de 8.87.

Es conveniente resaltar que tres de las áreas del examen tienen menos de 30 reactivos. Ese hecho disminuye la potencia de los indicadores por área, por lo cual, el cálculo se realizó para el total de reactivos del examen. Del mismo modo, es relevante destacar que es muy reducido el número de participantes en esta primera aplicación.



Figura 1. Frecuencia de reactivos por valor del índice de dificultad



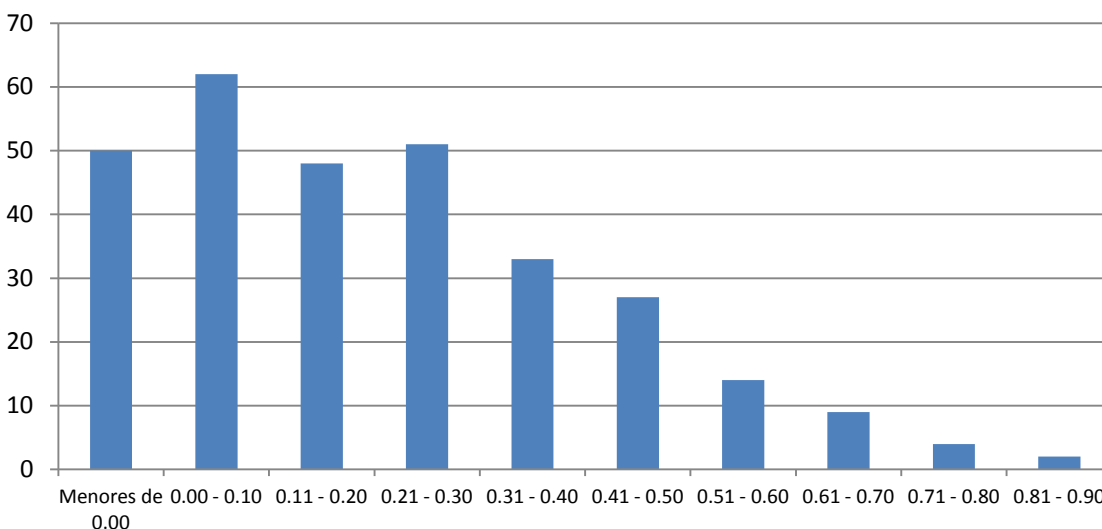
Los resultados mostrados en la figura 1 hacen suponer que el examen tiende a ser difícil, esto es, ligeramente más de la mitad de los reactivos tienen una dificultad menor o igual a 0.50, además la quinta parte de reactivos tiene un valor menor de 0.31. Este hecho se confirma porque se calculó la frecuencia de reactivos con dificultad menor o igual a 0.27, lo cual representa el 16%; en tanto solo el 10% de los reactivos tuvieron un valor de dificultad mayor o igual de 0.73.



EVALUACIÓN DEBATE 2014



Figura 2. Frecuencia de reactivos por valor del índice de discriminación



Con estos datos es importante destacar que uno de cada cinco reactivos puede conservarse, es decir 56 de ellos son excelentes, en tanto 35 son buenos. Aun así conviene revisar estos últimos con el fin de mejorarlos. En la misma proporción de reactivos que se quedarán tal como fueron diseñados, deben ser revisados a profundidad o descartarse, en tanto que una cantidad similar de reactivos debe descartarse de manera definitiva.

Con el objetivo de identificar si las variables asociadas con el éxito de los participantes cuando se ha aplicado la versión para el plan 90: promedio, año de egreso y actividad que desempeñan, se hicieron algunos análisis específicos. Se calculó la correlación de Pearson entre la calificación y el promedio de egreso. Se encontró una $r = 0.76$, $p < 0.5$. Esto indica una asociación fuerte entre ambas variables. En relación con el año de egreso hay un dato relevante, de los 42 aprobados, solo uno de ellos es egresado del plan 90. Esto indica que quienes tenían un año de haber concluido sus estudios tuvieron mayor éxito. Debido a que la mayoría de los participantes eran recién egresados, no se pudo realizar análisis para la variable actividad que desempeñan.

Falta aún considerar dos aspectos antes de tomar la decisión acerca de los reactivos, sobre todo aquellos con valores bajos en uno de los indicadores. Un primer momento es considerar los dos indicadores de manera simultánea para cada uno de los reactivos. De acuerdo con nuestra experiencia, si un reactivo tiene dificultad baja y discriminación alta, lo recomendable es revisarlo a la luz de cómo se distribuyeron las respuestas en los distractores. En contra parte, si se obtuvo dificultad alta y discriminación baja, debe considerarse que la reducida discriminación



se debe a la facilidad del reactivo. Esto es, si la mayoría de los participantes la respondieron correcta, el reactivo no es un buen reactivo para diferenciar entre el aspirante con buena ejecución de un aspirante con calificación baja.

Posterior al análisis anterior, es necesario realizar el análisis de la distribución de respuestas en los distractores. Esto es indispensable para los reactivos con dificultad buena (0.35 – 0.65) y con discriminación menor de 20. Es probable que una de las opciones falsas contenga parte de la información correcta, o bien que la base se plantee de modo que los mejores aspirantes se confunden y eligen esa opción por ser “atractiva”.

Consideraciones finales

No obstante que el examen tiende a ser difícil, los valores de los indicadores psicométricos fueron muy similares a los obtenidos en las trece aplicaciones que se han llevado a cabo de la versión EGC plan 90. En este sentido, el examen es un instrumento adecuado para identificar a los egresados capaces de manejar los contenidos curriculares en los que fueron formados, así como para aplicarlos en diferentes problemáticas educativas.

Se encontró una enorme diferencia entre los resultados de los egresados formados con el plan 90 o el plan 09; es evidente el éxito para los segundos (85% de participantes aprobados), y del mismo modo el fracaso de los primeros (2% de aprobación). Este dato no es indicador que el examen sea exclusivo para los egresados del plan 09. En nuestra opinión y a reserva de hacer análisis específicos, son las características de los participantes del plan 90 las que podrían explicar este hecho. Entre otras, el periodo tan largo entre su egreso y la aplicación, el promedio es menor al de la media de los egresados de ese plan y, finalmente, haber participado antes en el EGC plan 90.

Los resultados mostrados en el presente deben ser tomados con moderación, son varios los hechos que inciden en esta consideración. En primera instancia el reducido número de participantes lo cual reduce la potencia de los indicadores. Hay un aspecto medular que produjo una distribución demasiado irregular en las respuestas, nos referimos a que el 37% de los egresados que aplicaron el examen fueron formados con el plan 90 de la licenciatura. Esto significa que han participado en al menos tres ocasiones en el examen para ese plan sin haber alcanzado el título profesional. No era el objetivo del presente, pero resulta interesante realizar análisis de los reactivos diferenciados, es decir, calcular al menos el índice de dificultad solo para los egresados del plan 90. Con certeza se encontrarán valores distintos; no obstante esto llevará a reducir aún más la potencia de esta medida.



EVALUACIÓN DEBATE 2014



Referencias

- Backhoff, E., Sánchez, A., Peón, M., Monroy, L., y Tanamachi, L. (2006). Diseño y desarrollo de los exámenes de calidad y el logro educativos. *Revista Mexicana de Investigación Educativa*, 11(29), 617-638.
- Backhoff, E.; Larrazolo, y Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de habilidades y conocimientos básicos (EXHCOBA). *Revista electrónica de investigación educativa*, 2(1), 11-29.
- Baghaei, P. y Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple choice items. *Psychological Test and Measurement Modeling*, 53(2), 192-211.
- Bruno, J. y Dirkwager A. (1995). Determining the optimal number of alternatives to a multiple-choice test item: an information theoretic perspective. *Educational and Psychological Measurement*, 55(6), 959-966.
- Case, S. y Swanson, D. (2005). *Cómo construir preguntas de selección múltiple para ciencias básicas y ciencias clínicas*. Philadelphia: National Board of Medical Examiners.
- CENEVAL (2000). *Instructivo para la elaboración de reactivos*. EGEL Pedagogía-Ciencias de la educación. México: CENEVAL.
- CENEVAL. (2011). *Manual de elaboración de reactivos*. México: CENEVAL.
- CIEES, Comité de Artes, educación y humanidades. (2006). *Informe de evaluación diagnóstica. Licenciatura en psicología educativa*, Unidad Ajusco. México: Universidad Pedagógica Nacional. Documento inédito.
- Downing, S. M. (2006). Twelve steps for effective test development. En: S M. Downing, S.M. y T.M. Haladyna. *Handbook of test development*. New Jersey: Lawrence Erlbaum Associates.
- González, I.; García, J.; Pérez, C. y Soriano, R. (2011). *Programa educativo de la Licenciatura en psicología educativa. Instructivo de titulación*. México: UPN.
- Haladyna, T. y Downing, S. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement*, 53, 999-1010.
- Haladyna, T.; Downing, S. y Rodrigues, M. (2002). A review of multiple-choice item-writing for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Moreno, R.; Martínez, R. y Muñoz (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16(3), 490-407.
- Pérez, C.; Hernández, J.; Álvarez, Y. y Rubio, M. (2011). *Informe del Examen general de conocimientos de la Licenciatura en psicología educativa*. Documento interno. México: UPN.
- Rodrigues, M. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, Summer, 3-13.