



ISSN: 2448-6574

Evaluación del aprendizaje de contenidos del curso Psicología Evolutiva de la Aduldez y Vejez mediante una prueba objetiva

María Imelda González Mecalco
Alba Yanalte Álvarez Mejía
María Yolanda Quiroz Arce
Cuauhtémoc Gerardo Pérez López

mgonzalez@upn.mx
yalvarez@upn.mx
maquiroz@upn.mx
cgperez@upn.mx

Resumen

Con el fin de contar con un instrumento que permita hacer evaluación formativa de los estudiantes de la licenciatura, se aplicó por segunda ocasión una prueba para evaluar los aprendizajes adquiridos después de cursar las asignaturas Psicología evolutiva de la Infancia y la Adolescencia. Participaron en el estudio 274 alumnos inscritos en el cuarto semestre de la Licenciatura en Psicología Educativa de la Universidad Pedagógica Nacional, quienes cursaron la materia Psicología Evolutiva de la Aduldez y la Vejez en el semestre anterior a la aplicación. La prueba consta de 40 reactivos, 18 de los cuales se utilizaron en la primera aplicación, 5 de los cuales se modificaron en la estructura o en las opciones de respuesta, el resto se diseñaron para sustituir los reactivos con indicadores psicométricos bajos o para completar la tabla de especificaciones. Una vez calificadas las respuestas de los estudiantes, se calcularon los índices de facilidad y discriminación para cada uno de los reactivos y la escala total; estos son mucho más altos que los de la primera aplicación. Lo anterior permite concluir que se tiene ahora una prueba para evaluar los aprendizajes de los estudiantes cuyo comportamiento psicométrico es consistente y válido.

Palabras clave Evaluación educativa, evaluación objetiva de aprendizajes, instrumentos de evaluación objetiva

INTRODUCCIÓN

La Licenciatura en Psicología Educativa en la Universidad Pedagógica Nacional se cursa a lo largo de ocho semestres y está conformada por tres Fases y siete Líneas Curriculares. Entre estas Líneas se encuentra *Desarrollo y Aprendizaje*, la cual tiene como propósito formativo la comprensión y análisis de los procesos de desarrollo (cognitivos, afectivo, relacionales) y de



ISSN: 2448-6574

aprendizaje, así como su relación con los cambios derivados de la participación en los contextos educativos.

La Línea se compone por siete asignaturas, de las cuales tres abordan el estudio del desarrollo y cuatro el área del aprendizaje. En cuanto a las primeras, estas se enfocan al estudio del desarrollo evolutivo de las personas: infancia, adolescencia, adultez y vejez. Es en el curso Psicología evolutiva de la adultez y vejez (PEAV) en el que se pretende generar un instrumento para la evaluación de los aprendizajes, por lo que el presente documento es el resultado del estudio piloto del instrumento elaborado.

El estudio forma parte de un trabajo de investigación más amplio (que incluye la evaluación de aprendizajes del curso de *Psicología evolutiva de la infancia y Psicología evolutiva de la adolescencia*), en proceso hasta este momento. Para esta primera etapa del trabajo, misma que se reporta en el presente documento, se tomó una muestra de estudiantes de cuarto semestre de la carrera de Psicología Educativa, quienes cursaron la materia evaluada en tercer semestre. El trabajo tiene como propósito reportar el comportamiento del instrumento para evaluar los aprendizajes de los estudiantes de Psicología Educativa sobre contenidos de desarrollo evolutivo de la adultez y vejez, incluidos en el programa de estudio.

REFERENTES TEÓRICOS

La evaluación educativa puede ser vista desde distintas perspectivas teóricas, mismas que le asignan funciones específicas. Ketele & Roegiers (citado en Lukas & Santiago, 2009) hacen referencia a funciones tales como regulación, predicción, descripción, entre otras.

Desde hace varios años, Scriven (1967) formuló las funciones de la evaluación como sumativa y formativa. La primera con un carácter más bien de control y certificación de los resultados, en tanto que la formativa como medio para mejorar el objeto de evaluación. La evaluación sumativa resulta útil en algún tipo de estudio para realizar una valoración sobre un proceso concluido.

Mateo (2005) afirma que existen dos aproximaciones evaluativas: la evaluación centrada en los resultados y aquella orientada al estudio de los procesos, en la primera se hace uso de



ISSN: 2448-6574

recopilación de información que permite análisis fragmentarios y parciales, la segunda permite una aproximación global a la información.

Este trabajo pretende acercarse a una evaluación de carácter formativa, siendo coincidente con el planteamiento de Mateo (2005) cuando afirma que dependiendo del momento o de la necesidad del proceso de evaluación, se opta por una u otra función de la misma. De tal manera que la investigación de este reporte parcial informa de un acercamiento a una evaluación formativa, sin embargo, en esta primera etapa se da cuenta solamente del comportamiento de los reactivos de la prueba objetiva.

Para llevar a cabo la evaluación de los aprendizajes del curso ya mencionado, se revisaron los diversos procedimientos existentes para la recogida de información. Lukas & Santiago (2009) presentan una clasificación que da cuenta de aquellas técnicas de corte cuantitativo, así como de otras cualitativas. Este trabajo de investigación pretende acercarse a una evaluación de corte cualitativo, sin embargo, en su primera etapa se aplica una prueba objetiva que permite dar un panorama del aprendizaje sobre los contenidos de adultez y vejez por los estudiantes de Psicología Educativa con la finalidad de obtener información que permita el análisis cualitativo.

Se parte de la consideración de que las pruebas que se diseñan con la finalidad de evaluar los aprendizajes de los estudiantes deben ser estandarizadas y validadas, no solo por el impacto que algunas de estas tienen sobre los resultados de los estudiantes sino porque la toma de decisiones debe estar fundamentada.

Para el análisis de la prueba diseñada se consideró la facilidad/dificultad del reactivo y la discriminación del mismo. La dificultad se entiende como la proporción de personas que responden correctamente a un reactivo, de tal suerte que entre mayor sea la proporción, menor es su dificultad. Con la discriminación se espera que quienes tuvieron una puntuación alta en toda la prueba tengan más probabilidades de contestar correctamente al reactivo y que ocurra a la inversa con el estudiante que obtuvo puntuaciones bajas (Backhoff, Larrazolo & Rosas, 2000).

Un puntaje entre 0.28 y 0.72 se considera como un índice de facilidad apropiado, superior a esto se considera un reactivo que la mayoría responde como correcto, por lo que se sugiere revisar las opciones y la estructura del reactivo para reducir su nivel de facilidad. Por el contrario



ISSN: 2448-6574

un reactivo con un índice inferior a 0.28 tiende a provocar que pocos estudiantes lo respondan correctamente por lo que también será necesario revisar las opciones de respuesta, ya que existe la posibilidad de que alguna de ellas sea parcialmente correcta o que su estructura sea confusa para el estudiante, con ello se pretende incrementar su nivel de facilidad.

A fin de mejorar el instrumento de evaluación se retoman las recomendaciones de Ebel & Frisbie (citado en Backhoff, Larrazolo & Rosas, 2000) derivadas del puntaje de discriminación de los reactivos: si un reactivo obtiene un índice de discriminación superior a 0.39 se considera un reactivo de excelente calidad, por lo tanto se conserva; de 0.30 a 0.39 buena, esto implica que se puede mejorar; de 0.20 a 0.29 regular, en este caso es necesaria su revisión; de 0.00 a 0.19 pobre, estos se pueden descartar o revisar en profundidad; menor a 0.00 pésima, razón suficiente para descartarlos de la base de reactivos.

El presente trabajo tiene como propósito informar el comportamiento psicométrico, después de haber aplicado en dos ocasiones el instrumento para evaluar los aprendizajes de los estudiantes de Psicología Educativa sobre contenidos de desarrollo evolutivo de la adultez y vejez, incluidos en el programa de estudio.

MÉTODO

Participantes

En la primera aplicación se trabajó con una muestra de 258 estudiantes de cuarto semestre de la carrera en Psicología Educativa, de los cuales 131 pertenecen al turno matutino y 127 al vespertino. En la segunda, participaron 274 estudiantes, 172 del matutino, el resto del vespertino. Todos ellos inscritos en la carrera al momento de responder la prueba y habían cursado el semestre previo la materia PEAV.

Instrumento

En 2014 se diseñó una prueba objetiva con base en la revisión del *Programa del curso* y el *Documento técnico para la elaboración de los reactivos del Examen General de Conocimientos*. La descripción del procedimiento se puede consultar en (González, Álvarez & Nashiki, 2015).



ISSN: 2448-6574

La primera versión de la prueba se conformó con 41 reactivos, organizados por unidad como se muestra en la siguiente tabla.

Tabla 1. Temas y conceptos del curso PEAV

Unidad	Tema	Concepto
Unidad 1	Precurrentes básicos	Desarrollo, cambio y aprendizaje Envejecimiento exitoso
	Edad y relación con adultez y vejez	Edad Yo intemporal
	Modelos explicativos	Mecanicista, contextual, Ciclo vital, Ecológico
	Influencia de factores socioculturales	Influencia no normativa
Unidad 2	Cambio en los sistemas corporales	Capacidad de reserva
	Enfermedades, prevención y mantenimiento de la salud	Estilo de vida Enfermedad y malestar Salud
Unidad 3	Contexto familiar (estructura, matrimonio, divorcio, viudedad, casarse de nuevo, rol del abuelo)	Crisis de la mitad de la vida Divorcio Estructura familiar Rol del abuelo Duelo, pena, luto
	Contexto laboral (motivación y significado del trabajo, acontecimientos asociados a la productividad y no productividad, despido, jubilación)	Jubilación
Unidad 4	Memoria y atención	Memoria
	Procesos superiores y aprendizaje (pensamiento postformal, inteligencia, creatividad y sabiduría)	Inteligencia Sabiduría

Procedimiento

La primera fase del presente trabajo consistió en realizar el análisis de la información mediante la elaboración de una base de datos y un análisis de la discriminación y dificultad de cada reactivo, de la primera aplicación; dicha información se entregó a los participantes para su discusión en colegiado y toma de acuerdos sobre acciones a realizar a partir de los datos obtenidos.



ISSN: 2448-6574

En la segunda fase la tarea se centró en dos acciones: a) discutir la relevancia y pertinencia de los contenidos propuestos en la tabla de especificaciones y, b) decidir cómo proceder con cada uno de los reactivos a partir del análisis psicométrico de la prueba. De la primera acción se derivó la actualización de la tabla de especificaciones y se procuró mayor equilibrio en cuanto a la cantidad de contenidos evaluados con la prueba. Con la segunda, se definió cuáles reactivos debían eliminarse, cuáles podrían trabajarse en aras de mejorar su calidad y, por ende, sus indicadores y, cuántos debían ser diseñados para sustituir los reactivos eliminados y cubrir las condiciones de la tabla de especificaciones. Así, la segunda versión de la prueba consta de 40 reactivos, uno menos que la anterior.

Debe mencionarse que en ambas ocasiones, el instrumento se aplicó en las aulas de clases, en horario escolar y de manera simultánea. Esto implicó solicitar el apoyo a los docentes de 4º semestre para la aplicación a sus estudiantes.

RESULTADOS

De acuerdo con el objetivo del presente, para el análisis de los datos se calculó el índice de facilidad/dificultad y el de discriminación por cada uno de los reactivos de la segunda aplicación. Posterior a ello se organizaron en la tabla 2 a fin de a comparar los resultados de ambas aplicaciones. Es necesario aclarar que los números de la izquierda no indica con el número de reactivo, por lo cual, no hay correspondencia biunívoca entre los reactivos y el contenido que se evalúa. La tabla se construye así solo para visualizar, de manera general, que la segunda aplicación tiene mejores puntajes en los indicadores. En síntesis, la prueba es más fácil, pero sobre todo, discrimina significativamente mejor, condición que se pretende en todo instrumento de este tipo.

Tabla 2. Índice de facilidad y discriminación por año de aplicación

	Aplicación 2015		Aplicación 2016	
	Facilidad	Discriminación	Facilidad	Discriminación
1	0.95	0.06	0.29	0.31
2	0.27	0.21	0.45	0.31
3	0.59	0.37	0.65	0.50
4	0.5	0.24	0.16	0.15
5	0.55	0.19	0.42	0.27
6	0.62	0.17	0.52	0.43
7	0.28	0.30	0.70	0.50

8	0.44	0.20	0.84	0.35
9	0.45	0.27	0.68	0.47
10	0.52	0.34	0.41	0.20
11	0.61	0.20	0.88	0.34
12	0.91	0.11	0.55	0.42
13	0.84	0.20	0.70	0.43
14	0.71	0.21	0.91	0.39
15	0.52	0.39	0.24	0.35
16	0.84	0.13	0.87	0.38
17	0.34	0.23	0.50	0.45
18	0.16	0.16	0.61	0.38
19	0.49	-0.07	0.58	0.23
20	0.7	0.29	0.38	0.20
21	0.36	0.33	0.27	0.28
22	0.52	0.29	0.68	0.34
23	0.47	0.31	0.43	0.18
24	0.41	0.23	0.51	0.32
25	0.17	0	0.85	0.45
26	0.35	0.21	0.58	0.32
27	0.22	0.13	0.43	0.38
28	0.33	0.14	0.59	0.41
29	0.28	0.11	0.63	0.61
30	0.72	0.29	0.12	0.15
31	0.56	0.27	0.72	0.47
32	0.34	0.06	0.29	0.30
33	0.21	0.26	0.79	0.46
34	0.48	0.17	0.50	0.55
35	0.45	0.24	0.50	0.38
36	0.53	0.14	0.37	0.22
37	0.34	0.30	0.68	0.42
38	0.79	0.21	0.21	0.28
39	0.27	0.27	0.47	0.41
40	0.85	0.19	0.49	0.50
41	0.55	0.34		
Total	0.50	0.21	0.54	0.36

En la aplicación de 2015 se encontró que ocho reactivos tenían un índice de discriminación de excelente y buena calidad, es decir, podrían considerarse reactivos que diferencian a quienes saben la respuesta de quienes no la saben. De nivel de discriminación regular se tuvieron 18 reactivos y de calidad pobre resultaron según la ponderación un total de 15 reactivos (uno con discriminación negativa), para estos se acordó revisar su relación con el contenido a evaluar, su base y sus distractores.



ISSN: 2448-6574

De acuerdo con los puntajes de facilidad se observaron cinco reactivos en los que su nivel de facilidad es bajo y seis reactivos en los que su nivel de facilidad es alto; para ambos casos se acordó hacer una revisión de la estructura del reactivo, así como del comportamiento de sus opciones de respuesta para identificar qué aspecto modificar para propiciar un mayor indicador de identificación del aprendizaje de los estudiantes.

Posterior al análisis de facilidad y de discriminación se procedió a realizar un análisis del comportamiento de las opciones de respuesta de todos los reactivos. Así, se detectaron los reactivos cuya respuesta correcta es muy obvia, aquellos cuya opción correcta es difícil de encontrar por los estudiantes y el caso de los reactivos en los que una opción distractora pareciera ser la correcta puesto que la mayoría la elige cuando es errónea.

Este fue el conjunto de evidencias para realizar la siguiente acción -que se informa en el presente documento-, es decir, revisar, adecuar o eliminar los reactivos y, cuando fuera necesario, diseñar nuevos. De este modo, se debe mencionar que de los 41 reactivos de la primera aplicación, se decidió considerar solo 18 en la segunda versión. Debido a sus valores de facilidad y discriminación, 13 de estos se utilizaron tal como fueron aplicados en la primera versión; 5 fueron modificados principalmente en la base del reactivo, se procuró hacerla más clara y de menor tamaño.

En la tabla 3 se muestra el comportamiento psicométrico de estos 18 reactivos en ambas aplicaciones. Primero debe mencionarse que, en general, los reactivos fueron más fáciles; sin duda un dato con mayor relevancia es que en ninguno de estos reactivos el índice de discriminación se redujo y en solo uno este índice se incrementó una centésima; en el resto la discriminación se incrementó al menos cinco centésimas, resalta que en algunos el aumento fue mayor a 0.20 puntos.

La facilidad encontrada en los reactivos puede derivar de una adecuación general de la prueba, es decir, que el comportamiento del reactivo depende también del conjunto de reactivos en el que se encuentra. De igual manera, se considera que los docentes al frente de este curso y el resto de los de Psicología evolutiva han estado trabajando de manera colegiada para mejorar su práctica docente compartiendo propuestas didácticas.

Dado el interés del presente informe, mención especial se hace de los 5 reactivos que fueron objeto de modificación en su diseño. Así, el reactivo G es en el que se nota mayor crecimiento de ambos indicadores; con los ajustes realizados en el mismo se convirtió en un reactivo con muy buena discriminación. De manera similar se comportaron los reactivos Q y O; aunque este último mantuvo la tendencia a ser reactivo fácil. En tanto en el reactivo M se incrementó favorablemente la facilidad y se obtuvo una discriminación buena. Por último, en el reactivo L se mantuvo la facilidad y se mejoró la discriminación.

Tabla 3. Índices de facilidad y discriminación por aplicación y reactivo

	Aplicación 2015		Aplicación 2016	
	Facilidad	Discriminación	Facilidad	Discriminación
A	0.27	0.21	0.29	0.31
B	0.55	0.19	0.45	0.31
C	0.52	0.34	0.65	0.50
D	0.44	0.20	0.42	0.27
E	0.59	0.37	0.52	0.43
F	0.62	0.17	0.70	0.50
G*	0.17	0	0.68	0.47
H	0.61	0.20	0.55	0.42
I	0.91	0.11	0.91	0.39
J	0.84	0.20	0.70	0.43
K	0.49	-0.07	0.24	0.35
L*	0.56	0.27	0.51	0.32
M*	0.22	0.13	0.58	0.32
N	0.21	0.26	0.12	0.15
O*	0.85	0.19	0.79	0.46
P	0.55	0.34	0.50	0.55
Q*	0.53	0.14	0.68	0.42
R	0.27	0.27	0.21	0.28

Como se dijo ya, la prueba tiene 40 reactivos, en consecuencia se diseñaron 22 reactivos nuevos; esto para sustituir los reactivos que por los valores de sus indicadores se eliminaron, o bien, para cubrir totalmente los temas en la tabla de especificaciones.

CONCLUSIONES

El objetivo de contar una segunda versión de la prueba que evaluara los contenidos significativos del programa de la asignatura PEAV, además de tener mejores indicadores



ISSN: 2448-6574

psicométricos fue lograda en su cabalidad. De acuerdo con los hallazgos antes descritos, es posible decir que la prueba es un instrumento de evaluación objetiva que podrá ser utilizado para medir el aprendizaje de los contenidos. Por supuesto, esto deberá ser corroborado en aplicaciones posteriores.

Los criterios de discriminación y facilidad permitieron identificar de manera efectiva el comportamiento de los reactivos en particular y del examen en general. Con ello, se da pauta para la creación de instrumentos con un grado de validez aceptable, al reconocer los aspectos de mejora.

La aplicación de este instrumento, a una muestra y el análisis de su comportamiento dan pie al desarrollo de instrumentos similares con otros cursos de la Línea Curricular Desarrollo y Aprendizaje, de tal manera que a largo plazo pueda contarse con instrumentos probados para evaluar los aprendizajes que se construyen en los cursos de la línea curricular.

La construcción de este tipo de instrumentos permite avanzar hacia la evaluación formativa y objetiva que dé cuenta de los aprendizajes de los estudiantes de psicología educativa.

Para continuar con este tipo de evaluaciones es necesario realizar un análisis detallado del contenido de los reactivos y de su funcionamiento, para garantizar que el examen incluya los contenidos esenciales y se muestre equilibrado en función del currículo.

REFERENCIAS

- Backhoff, E., Larrazolo, N. & 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). Consultado el 20 de junio de 2014 en: <http://redie.uabc.mx/index.php/redie/article/viewFile/15/26>
- Baghaei, P. & 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. & 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.



ISSN: 2448-6574

- Haladyna, T. & 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. & Rodríguez, M. (2002). A review of multiple-choice item-writing for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Lukas & Santiago (2009). *Evaluación educativa*. Madrid: Alianza.
- Mateo, J. (2005). *La evaluación educativa, su práctica y otras metáforas*. México: Alfaomega.
- Moreno, R., Martínez, R. & Muñoz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema* 16(3), 490-497. Consultado el 11 de marzo de 2014 en: <http://www.psicothema.com/pdf/3023.pdf>
- Pérez.C. G., Bollás, P., González, M. I, Álvarez, A. Y. & Castro, R. (2014). Examen general de conocimientos para la licenciatura en psicología educativa, plan 09, Universidad Pedagógica Nacional. Indicadores psicométricos. Memorias Congreso Internacional de Educación, Evaluación.
- Rodríguez, 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.
- Scriven, M. (1967). The methodology of evaluation. En R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally.