

VI Congreso Internacional de Investigación y Práctica Profesional en Psicología
XXI Jornadas de Investigación Décimo Encuentro de Investigadores en
Psicología del MERCOSUR. Facultad de Psicología - Universidad de Buenos
Aires, Buenos Aires, 2014.

Aplicación de la teoría clásica de test a la evaluación del rendimiento en estudiantes universitarios.

Alvarez Ponte, Lucía Inés y Galibert, María
Silvia.

Cita:

Alvarez Ponte, Lucía Inés y Galibert, María Silvia (2014). *Aplicación de la teoría clásica de test a la evaluación del rendimiento en estudiantes universitarios. VI Congreso Internacional de Investigación y Práctica Profesional en Psicología XXI Jornadas de Investigación Décimo Encuentro de Investigadores en Psicología del MERCOSUR. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-035/792>

ARK: <https://n2t.net/ark:/13683/ecXM/7D2>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

APLICACIÓN DE LA TEORÍA CLÁSICA DE TEST A LA EVALUACIÓN DEL RENDIMIENTO EN ESTUDIANTES UNIVERSITARIOS

Alvarez Ponte, Lucía Inés; Galibert, María Silvia

UBACyT, Universidad de Buenos Aires - ANPCyT, Agencia Nacional de Promoción Científica y Tecnológica

RESUMEN

La evaluación universitaria ha sido objeto de abundante investigación en los últimos años (Arribas, 2012; Buscá et al, 2011). Aunque dicha evaluación tiene la doble función de formativa y certificadora (Villardón, 2006; Taras, 2005; López, 2009) cuando se alude al rendimiento académico se piensa simplemente en las calificaciones (Rodríguez y Ruiz, 2011), función certificadora. Luego, la calificación tiene implicancias relevantes para el evaluado y condiciona todo el proceso de enseñanza - aprendizaje (Biggs, 2005; Sans, 2005; Cabaní y Carretero, 2003). Por tanto la evaluación debe planificarse y los exámenes ser debidamente confeccionados. En este trabajo se evalúan las propiedades psicométricas de un examen de probabilidad aplicando la Teoría Clásica de Tests para mejorar la calidad de la evaluación. El examen consistió en 5 problemas de elaboración desglosados en 10 subítems. Fue administrado a 34 estudiantes de ingeniería (CABA). Cuatro problemas resultaron muy difíciles. La discriminación (correlación ítem-test) fue baja para dos problemas y su exclusión aumentaría la confiabilidad α de Cronbach de .581 a .686, aceptable para un test corto. El análisis psicométrico del examen permitió detectar los ítems que afectan la evaluación en su función certificadora y orientar en qué temas focalizar para mejorar el rendimiento (función formativa).

Palabras clave

Evaluación del rendimiento, Examen, Estudiantes universitarios, Confiabilidad, Psicometría, TCT

ABSTRACT

APPLICATION OF CLASSICAL TEST THEORY TO THE ACHIEVEMENT ASSESSMENT IN UNIVERSITY STUDENTS

University assessment has been the subject of much research in recent years (Arribas, 2012; Buscá et al, 2011). Although this assessment has the double function of training and certification (Villardón, 2006; Taras, 2005; López, 2009), when people refer to achievement assessment they just think on scores (Rodríguez & Ruiz, 2011), which is mainly connected to the certification role. Thus, scoring has important implications for the examinees and this conditions all the process of teaching - learning (Biggs, 2005; Sans, 2005; Cabani & Carter, 2003). Therefore, the assessment should be planned and exams should be carefully made. In this study, psychometric properties of a probability exam are evaluated by using Classical Test Theory in order to improve the quality of the assessment. The exam consisted of 5 problems broken down into 10 subitems. It was administered to 34 engineering students (CABA). Four problems were very difficult. Discrimination (item-test correlation) was low for two problems and its exclusion would improve the Cronbach's index reliability from $\alpha = .581$ up to $\alpha = .686$, acceptable for a short test. The psychometric analysis allowed to detect those items that affect the assessment in its certification function and the topics

that are necessary to focus on, to improve student achievements (formative function).

Key words

Achievement assessment, Exam, University students, Reliability, Psychometry, Classical Test Theory

La evaluación universitaria ha sido objeto de abundante investigación en los últimos años (Arribas, 2012; Buscá et al, 2011). La bibliografía remite a dos funciones fundamentales y complementarias: la formativa y la certificadora (Villardón, 2006; Taras, 2005, López, 2009). La función formativa focaliza en el proceso de aprendizaje y en el protagonismo del alumno en dicho proceso. En relación a este aspecto, hay quienes propician una mayor participación del alumno en su evaluación (Goñi, 2005; Falchikov, 2005; Escudero, 2010; López, 2009; Bretones, 2008, Pérez, Julián y López, 2009). Pero, según señalan Bordas y Cabrera (2001), todavía la práctica real está lejos de estas propuestas teóricas. De ahí que, en general, cuando se alude al rendimiento académico se piense simplemente en las calificaciones (Rodríguez y Ruiz, 2011), lo que se vincula más con la función certificadora. Luego, la calificación tiene implicancias relevantes para el evaluado y condiciona el proceso de enseñanza - aprendizaje (Biggs, 2005; Sans, 2005; Cabaní y Carretero, 2003). Por eso es importante que la evaluación sea planificada y los instrumentos de medición, los exámenes, debidamente confeccionados.

La Teoría Clásica de Tests (TCT) y la Teoría de Respuesta al Ítem (TRI) proporcionan modelos que orientan la construcción de pruebas con propiedades psicométricas adecuadas. La TRI tiene muchas ventajas teóricas con respecto a la TCT pero ciertas limitaciones para sus aplicaciones, por lo que no está tan difundida a nivel del aula (Muñiz, 2002; Martínez Arias, R., 2006, Santisteban Requena, 2009; Attorresi et al 2009; Abal et al 2010). En efecto, la mayor complejidad de sus modelos demanda mayor especialización o formación matemática de los evaluadores, disponibilidad de softwares específicos y, lo que es un obstáculo muchas veces insuperable, tamaños de muestras muy grandes que exceden la realidad corriente de las aulas. Por estas razones la TCT sigue siendo muy utilizada en la construcción de pruebas psicológicas, generalmente de opción múltiple, aunque no tanto a nivel del aula, principalmente por desconocimiento. Sin embargo, con la debida difusión podría implementarse en el ámbito educativo ya que resultaría accesible aun a docentes no expertos en Psicometría, quienes podrían aplicar algunos criterios derivados de ella para mejorar la calidad de sus exámenes. Este trabajo tiene como objetivo evaluar las propiedades psicométricas de un examen de probabilidad y de los ítems que lo componen mediante el cálculo de los índices de la Teoría Clásica de Tests.

Método

Instrumento de Medición. Es un examen que consiste en 5 problemas de elaboración. El primero (P1) comprende dos subítems que evalúan respectivamente el conocimiento de la relación entre las distribuciones Exponencial - Poisson y la aplicación de la distribución Binomial condicional a una Poisson. Cada subítem vale 1 punto. El segundo problema (P2) evalúa las distribuciones Uniforme y Normal con aplicación del Teorema Central del Límite. Vale 2 puntos. El tercer problema (P3) consta de dos subítems, de un punto cada uno, que requieren conocimientos de la distribución Hipergeométrica y de Probabilidad Condicional. El cuarto problema (P4) vale 2 puntos y corresponde a hallar la función de probabilidad de una variable aleatoria que está condicionada a una Binomial. El quinto problema (P5) involucra a la distribución conjunta de dos variables continuas y está desglosado en 4 subítems de medio punto cada uno. La asignación de puntajes a cada ítem podía llegar a tener una precisión de hasta 12,5 centésimos. La calificación total en el examen se obtuvo como la suma de los puntajes en cada subítem, por lo que resulta en la escala habitual de calificaciones de 0 a 10.

Participantes. Participaron 34 alumnos de segundo año de Ingeniería Industrial de una universidad privada de la Ciudad Autónoma de Buenos Aires.

Análisis psicométrico. Se calcularon los índices psicométricos de dos maneras: considerando como unidad cada uno de los 5 problemas y también tomando como unidad cada uno de los 10 subítems en los cuales dichos problemas se desglosaban. Se obtuvieron para ambos casos el coeficiente alpha de Cronbach y los índices característicos de los ítems. Dichos índices son la media y desviación estándar del puntaje en cada ítem, su índice de dificultad calculado como el cociente entre la media y el valor total del ítem, el índice de confiabilidad, el índice de correlación lineal r de Pearson entre el ítem y el puntaje en el examen sin el ítem y el coeficiente de Cronbach que tendría el examen si se excluyera el ítem. Obsérvese que el índice de dificultad está comprendido entre 0 y 1, donde 0 corresponde a extrema dificultad y 1 a extrema facilidad. El coeficiente α de Cronbach se obtuvo para el total de ítems y luego se obtuvo para cada subconjunto de ítems conforme se iban excluyendo de manera sucesiva aquellos cuya presencia hacían disminuir este coeficiente. También se calcularon los resúmenes estadísticos de las calificaciones.

Resultados

Considerando el conjunto de 5 problemas se obtuvo un coeficiente $\alpha = .581$ que podría incrementarse hasta $.686$ excluyendo a los problemas 1 y 3. Para los 10 subítems $\alpha = .613$ y alcanzaría el valor $.666$ excluyendo a los subítems 1, 4 y 5, los cuales tuvieron las menores correlaciones ítem-test. Los índices de los ítems y los resúmenes de las calificaciones se exhiben en las tablas 1 y 2.

Tabla 1. Índices psicométricos de los subítems y de los problemas

SUBÍTEM PROBLEMA	MEDIA	DIFICULTAD MEDIA/ VALOR	DESVIÓ ESTÁNDAR	CONFIABILIDAD	COEFICIENTE r ÍTEM-TEST	ALPHA SIN EL ÍTEM
1	0,485	.4850	0,437	.121	.073	.633
2	0,088	.0880	0,249	.101	.299	.593
P1	0.574	.2870	0.544	.221	.163	.603
3	0,640	.3200	0,705	.485	.440	.542
P2	0.640	.3200	0.705	.485	.440	.466
4	0,765	.7650	0,397	.077	.007	.641
5	0,537	.5379	0,482	.207	.218	.603
P3	1.301	.6505	0.669	.284	.119	.635
6	0,757	.3785	0,856	.654	.488	.536
P4	0.757	.3785	0.856	.654	.488	.425
7	0,173	.3460	0,219	.128	.511	.571
8	0,125	.2500	0,203	.126	.557	.570
9	0,096	.1920	0,163	.108	.611	.574
10	0,099	.1980	0,186	.118	.574	.572
P5	0.493	.2465	0.663	.480	.512	.430

Tabla 2. Resúmenes estadísticos de las calificaciones.

MÍNIMO	MEDIA	MEDIANA	MÁXIMO	DESVIÓ ESTÁNDAR	ASIMETRÍA	CURTOSIS
1	3,7647	3,25	9,75	2,1238	0,9174	0,4265

Discusión

La confiabilidad de 0,581 para los 5 problemas es satisfactoria teniendo en cuenta la muy pequeña longitud del examen concebido como un test en el marco de la TCT. El índice de Spearman - Brown, si el examen se proyectara a uno de 15 ítems, predice un $\alpha = .81$, confiabilidad considerada alta para una prueba de rendimiento (Magnusson, 1982; Thorndike, 1989).

Casi todos los subítems y, por ende, los problemas, resultaron muy difíciles. En efecto, sólo el problema P3 fue medianamente fácil (.6505). Por esta misma razón, que lo diferencia de los demás, tuvo la menor correlación con el resto de la prueba, lo que afecta su confiabilidad en términos de consistencia interna. Por ello al excluirlo aumentaría la confiabilidad del examen. Otro problema que podría excluirse para aumentar la confiabilidad es P1 que resultó difícil por su segundo subítem, aunque afectó la confiabilidad por su parte más fácil, el primer subítem. Cuatro de los subítems difíciles (7, 8, 9 y 10) integran un único problema, P5, que evalúa la distribución conjunta de variables continuas. Estos subítems tuvieron buenos índices de correlación ítem-test pero hay que tener en cuenta que ésta podría estar inflada precisamente por la interrelación entre ellos. Sin embargo, al examinar su correlación como grupo con el resto (P5 vs los demás), sigue siendo alta (.512). También es alta su confiabilidad pero este problema es el más difícil. Los problemas P2 y P4, también difíciles, tuvieron como el P5 muy buenos índices de confiabilidad y correlación. Puede observarse, entonces, que hay una asociación entre la dificultad y la confiabilidad. Siendo la mayoría difíciles, son los problemas o subítems medianamente fáciles los que terminan afectando la consistencia interna. Si el objetivo fuera puramente psicométrico podrían simplemente excluirse dichos ítems del examen. Pero como se trata de una prueba de rendimiento, hay que evaluar si no se está dejando de representar una parte importante del contenido. En el presente caso, excluir el P3 dejaría sin evaluar la distribución Hipergeométrica. La homo-

geneidad no está siendo afectada por la no pertinencia de lo que se evalúa en este problema sino por su discrepancia en dificultad con el resto. Sería más razonable buscar la solución en reforzar desde la enseñanza los otros contenidos que resultaron tan difíciles o bien reformular los problemas que aluden a dichos contenidos de modo que resulten más fáciles. En su versión actual, este examen es más apto para la evaluación en su función formativa que certificadora. A este respecto Cano de Becerra (1971) aconseja que los puntajes tengan una cierta asimetría positiva para los exámenes administrados durante el período de la cursada y negativa en la instancia final. Esta sugerencia se fundamenta en el hecho de que una distribución asimétrica positiva discrimina mejor a los alumnos de altos rendimientos, lo cual es preferible durante el proceso de enseñanza - aprendizaje para estimularlos a empeñarse más; por el contrario, en el examen final conviene discriminar mejor a los de peor desempeño mediante una distribución asimétrica negativa de modo que se esté seguro de que desapueba quien realmente tiene un pobre rendimiento. Lamentablemente el presente examen tuvo una marcada asimetría positiva, justamente en virtud de su gran dificultad (el puntaje promedio 3,76 está por debajo de la nota de aprobación), y fue administrado en una instancia final; es decir, en función primordialmente certificadora. Su análisis psicométrico permitió detectar que no es tan apropiado para la instancia final pero sí para la evaluación durante el proceso de enseñanza-aprendizaje y que debería reforzarse mejor la enseñanza de los temas que resultaron particularmente difíciles a saber: Teorema Central del Límite, función de probabilidad condicional y distribución conjunta entre variables continuas.

BIBLIOGRAFIA

- Arribas, J. (2012). El rendimiento académico en función del sistema de evaluación empleado. *RELIEVE*, 18,1,3. <http://www.uv.es/RELIEVE>
- Biggs, J. (2005). *Calidad del aprendizaje universitario*. Madrid: Narcea.
- Buscá, F., Cladellas, L., Calvo, J., Martín, M., Padrós, M. y Capllonch, M. (2011). Evaluación formativa y participativa en docencia universitaria. Un estudio sobre los artículos publicados en revistas españolas entre 1999 y 2009. *Aula Abierta*, 39, 2, 137-148.
- Cano de Becerra, F. (1971). Elementos de Estadística al Servicio de la Evaluación del Rendimiento. *Revista de Psicología*, 16(1/2):61-77.
- Cavani, M. y Carretero, R. (2003). La promoción de estudiantes estratégicos a través del proceso de evaluación que proponen los profesores universitarios. En C. Monereo y J. L. Pozo, *La universidad ante la nueva cultura educativa: enseñar y aprender para la autonomía*, 173-190. Madrid: Síntesis.
- López, V. (2009). *La evaluación formativa y compartida en docencia universitaria: propuestas, técnicas, instrumentos y experiencias*. Madrid: Narcea.
- Magnusson, D. (1982). *Teoría de los Tests*. Ed. Trillas, México.
- Rodríguez, M. y Ruiz, M. (2011). Indicadores de rendimiento de estudiantes universitarios: calificaciones versus créditos acumulados. *Revista de Educación*, 355, 467-492.
- Sans, A. (2005). *La evaluación de los aprendizajes: construcción de instrumentos*. Cuadernos de docencia universitaria, 2. Barcelona: Octaedro-ICE.
- Taras, M. (2005). Assessment - sumative and formative- some theoretical reflections. *British Journal of Educational Studies*, 53, 4, 466-478.
- Thorndike, R. (1989). *Psicometría Aplicada*. Ed. Limusa, México.
- Villardón, L. (2006). Evaluación del aprendizaje para promover el desarrollo de competencias. *Educatio. Siglo XXI*, 24, 15-35.