

Utilización de Análisis Cuantitativos para el Estudio de un Sistema de Evaluación Universitario

RODOLFO UNGERFELD*

Recibido: 23/07/05

Aprovado: 14/10/05

* Facultad de Veterinaria, Montevideo, Uruguay, piub@internet.com.uy

Resumen: en el trabajo se presenta un análisis cuantitativo de los resultados obtenidos en las evaluaciones realizadas durante 3 años por estudiantes universitarios. Se analizan y discuten las posibles causas del resultado de cada evaluación. Por último, se plantea la utilidad de realizar análisis cuantitativos como forma de priorizar sobre qué instancias de evaluación debe trabajarse.

Palabras clave: evaluación, pruebas, evaluación curricular, técnicas de evaluación

The Utilization of quantitative analysis for the study of a higher education evaluation system.

Abstract: the article analyses quantitatively the results obtained in the evaluations performed during 3 years by university students. The possible causes of the results of each evaluation technique are analyzed and discussed. The importance of quantitative analyses of the student's results, and how it may be considered for possible changes on the mechanisms used for evaluation is also discussed.

Key-words: Evaluation; Contests; Curricular evaluation; Evaluation techniques

1. Introducción

Los sistemas de evaluación, además de permitir la acreditación de los estudiantes, deben servir como mecanismo permanente de retroalimentación de la enseñanza. Los cambios en las técnicas de evaluación utilizadas suelen ser trabajosos, ya que luego de la discusión y el análisis de los mismos es recomendable realizar estudios repetidos de resultados, ensayos previos en donde se determine la exactitud de los planteos respecto a las respuestas que se pretende, etc. En ese marco es importante generar información objetiva que permita analizar los resultados, para así centrar los esfuerzos en las áreas temáticas o en las actividades cuyo resultado lo demande prioritariamente. A su vez, el dictado de estas mismas áreas debería ser objeto de discusión y evaluación por parte del equipo docente a cargo.

Pese a la formación académica que se aspiraría por parte de los docentes universitarios, es poco frecuente que se analicen cuantitativamente los resultados de las evaluaciones. El objetivo del presente trabajo es discutir y analizar las ventajas de aplicar análisis cuantitativos de los resultados de las evaluaciones en cursos universitarios, así como la potencialidad de utilizarlo como insumo para la redefinición de los sistemas y técnicas de evaluación del curso. Para ello, se utilizan como ejemplo los resultados obtenidos en las instancias de evaluación de un curso curricular de la Facultad de Veterinaria.

2. Sistema de evaluación

El curso se imparte desde el Año 1999 en el tercer semestre del Plan 98 de la Facultad de Veterinaria. Hasta 1998 el curso tenía una duración anual, por lo que a partir de 1999 fue necesario readecuar los contenidos y la estructura del curso a la nueva realidad. Si bien los contenidos fueron rediscutidos y readecuados en el nuevo contexto, no se introdujeron mayores cambios en la forma de evaluación, aplicándose el mismo sistema de evaluación por parciales teóricos y seminario que se utilizaba en el curso anterior. La única instancia de evaluación que se incorporó fue la evaluación por parciales prácticos que no se utilizaba hasta el momento. Mientras que los parciales teóricos y prácticos son diseñados de común acuerdo entre los docentes involucrados, el criterio y los mecanismos de evaluación de los seminarios dependen fundamentalmente del docente responsable de cada uno.

La evaluación objetiva del estudiante se basa en el puntaje de dos parciales teóricos, un control globalizador práctico y el seminario, generando cada instancia un máximo posible de 25 puntos (100 en total). Para obtener la ganancia del curso el estudiante debe superar los 45 puntos, de los que al menos 30 deben obtenerse a través de los 3 parciales. Una vez superado ese umbral, y hasta los 64,9 puntos totales, el estudiante queda habilitado a rendir el examen común. Si el puntaje final está entre 65 y 74,9 el estudiante podrá rendir un examen calificado. Por último, si el estudiante obtiene más de 75 puntos incluyendo un mínimo de 55 obtenidos a través de los 3 parciales exonera el curso.

Los parciales teóricos son escritos, incluyendo una importante variedad de formas estructuradas. Se incluyen preguntas de múltiple opción, de verdadero/falso, de respuesta libre corta y problemas. El control globalizador práctico se basa en la proyección de transparencias de registros gráficos, tablas o montajes experimentales realizados durante el curso práctico, debiendo el estudiante responder una serie de preguntas estructuradas. El control es tomado en dos instancias, coincidentes temporalmente con los parciales teóricos. De acuerdo a la carga temática, el puntaje ha sido 12,5 y 12,5 en cada instancia (2001) y 10 y 15 en la primera y segunda instancia respectivamente (2002 y 2003). Un comentario que se reitera en las evaluaciones por

parte de muchos estudiantes es que este es un sistema que fomenta lo memorístico, ya que deben entender los registros, analizarlos e interpretarlos y responder en un lapso de tiempo muy corto.

El seminario, de acuerdo al Plan 98, es la actividad integradora en el semestre. Es realizado por un grupo de 6 a 10 estudiantes, y debe ser monitoreado por dos docentes. En la práctica, y dado el alto número de estudiantes, el monitoreo es realizado por un solo docente. Dado que es la actividad integradora del semestre, puede ser monitoreado por docentes de cualquiera de las Áreas académicas vinculadas al semestre (Fisiología, Inmunología, Microbiología, Zootecnia) a las que habitualmente –dada la insuficiencia de docentes- se agregan docentes de Áreas académicas ya cursadas por los estudiantes (Biofísica, Genética, Histología y Embriología). El puntaje asignado es individual para cada estudiante, siendo un máximo de 22 puntos adjudicado por el docente monitor. Al finalizar el semestre se realiza una jornada en que los seminarios son presentados en forma de póster, y cada póster es evaluado por parte de al menos 3 docentes monitores de otros seminarios, adjudicándose un puntaje común al grupo de hasta 3 puntos. El seminario permite incorporar actividades con perfil de investigación, las que son necesarias para potenciar en múltiples formas el desarrollo de la enseñanza (Compiano y Giarrizo, 1995; Ungerfeld, 2004). La evaluación de alternativas, balancear el peso de las evidencias, la interpretación crítica de publicaciones –y por tanto el manejo del estado de la información- son componentes esenciales en la construcción de los argumentos científicos (Driver et al., 2000).

Los sistemas de evaluación, además de permitir la acreditación de los estudiantes, deben servir como mecanismo permanente de retroalimentación de la enseñanza.

3. Materiales y métodos

Se analizaron los resultados finales de todos los estudiantes que completaron todas las instancias de evaluación (parciales teóricos, parciales prácticos y seminario) durante los años 2001, 2002 y 2003. Esta población se compone de 652 estudiantes, 187, 241 y 224 para los años 2001, 2002 y 2003 respectivamente.

Los datos fueron analizados de acuerdo al año y a la categoría del estudiante al finalizar el curso (Pérdida del curso, Examen común, Examen calificado y Exonerado), considerándose también cada instancia de parcial teórico desglosada. Para el análisis del resultado de los resultados del CGP se consideraron los años 2002 y 2003; para analizar los resultados de los seminarios solamente se consideró el año 2003.

Las frecuencias (estudiantes por categoría al finalizar el curso) fueron comparados por test de chi cuadrado, los resultados de parciales y de seminarios por ANOVA, y la relación entre el resultado del primer y el segundo parcial, y entre los resultados de los parciales teóricos y prácticos se relacionaron mediante una regresión lineal.

4. Resultados

Resultados finales por año

La cantidad de estudiantes en cada categoría fue diferente entre los años estudiados ($P=0,045$; Tabla 1). Este resultado global es generado por una diferencia entre los años 2001 y 2002 ($P=0,01$), mientras que la diferencia entre los años 2001 y 2003, y 2002 y 2003 no alcanzó significancia estadística ($P>0,05$).

Resultados de la evaluación por parciales teóricos por año

El resultado de la evaluación por parciales teóricos (puntaje máximo de 50 puntos) fue significativamente diferente de acuerdo al año ($P<0,0001$), siendo el del año 2002 ($23,3 \pm 0,3$) más bajo que el del año 2001 ($24,9 \pm 0,5$, $P=0,01$) y que el del año 2003 ($25,5 \pm 0,4$). La diferencia entre el año 2001 y el año 2003 no alcanzó significación estadística.

En una segunda instancia se desglosó el total en subgrupos de acuerdo al resultado del curso, determinándose como incidió cada subgrupo sobre la diferencia final por año (Tabla 2).

Tabla 1. Cantidad (%) de estudiantes que finalizaron el curso en cada categoría los años 2001, 2002 y 2003.

	2001 (%)	2002 (%)	2003 (%)
Pérdida del curso	33 (17,6)	34 (14,1)	34 (15,2)
Examen común	90 (48,1)	129 (53,5)	125 (55,8)
Examen calificado	43 (23,0)	69 (28,6)	50 (22,3)
Exonera	21 (11,2)	9 (3,7)	15 (6,7)
Total	187	241	224

Tabla 2. Promedio (\pm EE) del resultado de los parciales teóricos en función de la categoría final del estudiante durante los años 2001, 2002 y 2003 (puntaje máximo posible: 50).

	Pérdida de curso	Común	Calificado	Exonerado
2001	$14,9 \pm 0,7$	$23,3 \pm 0,4^a$	$30,7 \pm 0,4^a$	$35,6 \pm 1,3$
2002	$16,3 \pm 0,4$	$22,0 \pm 0,3^b$	$27,9 \pm 0,3^b$	$33,6 \pm 0,8$
2003	$16,1 \pm 0,5$	$24,4 \pm 0,3^c$	$31,0 \pm 0,5^a$	$37,6 \pm 0,6$

Diferentes letras en una misma columna: $P<0,001$

Resultados de cada evaluación parcial teórica

Cuando el resultado de los parciales se descompone entre el primer y el segundo parcial se observan diferencias significativas todos los años (Tabla 3, puntaje máximo por parcial: 25 puntos). Más allá de que las diferencias se invirtieron de acuerdo al año, en todos los casos se alcanzaron significancias con $P < 1 \times 10^{-15}$.

Tabla 3. Puntaje promedio (\pm EE) obtenido en cada parcial teórico durante los años 2001, 2002 y 2003 (puntaje máximo posible en cada parcial: 25).

	2001	2002	2003
Parcial 1	11,5 \pm 0,3 ^a	12,8 \pm 0,2 ^b	11,9 \pm 0,2 ^a
Parcial 2	13,4 \pm 0,3 ^a	10,5 \pm 0,2 ^b	13,5 \pm 0,2 ^a

Para una misma fila: ^a vs ^b; $P < 0,001$.

Tabla 4. Puntaje promedio (\pm EE) en cada parcial desglosado, de acuerdo a la categoría en que el estudiante finalizó el curso (2001; puntaje máximo por parcial: 25).

	Pérdida de curso	Examen común	Examen calificado	Exonera
Parcial 1	7,0 \pm 0,3 †	10,6 \pm 0,3 [*]	14,2 \pm 0,3 [*]	16,8 \pm 0,7 [*]
Parcial 2	7,8 \pm 0,4	12,7 \pm 0,3	16,4 \pm 0,3	18,8 \pm 0,7

† $P = 0,06$; ^{*} $P < 0,0001$

Tabla 5. Puntaje promedio (\pm EE) en cada parcial, desglosado de acuerdo a la categoría en que el estudiante finalizó el curso (2002; puntaje máximo por parcial: 25).

	Pérdida de curso	Examen común	Examen calificado	Exonera
Parcial 1	8,8 \pm 0,3 [*]	12,1 \pm 0,2 ^{**}	15,5 \pm 0,2 ^{**}	18,0 \pm 0,6 †
Parcial 2	7,5 \pm 0,3	9,9 \pm 0,2	12,4 \pm 0,2	15,6 \pm 0,8

^{*} $P < 0,01$; ^{**} $P < 0,0001$; † $P = 0,07$

El resultado del segundo parcial se correlacionó (para los 3 años: $P < 1 \times 10^{-18}$) con el resultado del primero. En la Figura 1 se presentan los resultados de los dos parciales para los años 2001 (Figura 1A; $r = 0,8$), 2002 (Figura 1B, $r = 0,5$) y 2003 (Figura 1C; $r = 0,7$).

Tabla 6. Puntaje promedio (\pm EE) en cada parcial, desglosado de acuerdo a la categoría en que el estudiante finalizó el curso (2003; puntaje máximo por parcial: 25).

	Pérdida de curso	Examen común	Examen calificado	Exonera
Parcial 1	7,2 \pm 0,4 [*]	11,4 \pm 0,2 ^{**}	14,7 \pm 0,3 ^{**}	17,9 \pm 0,5 ^{***}
Parcial 2	8,9 \pm 0,3	12,9 \pm 0,2	16,3 \pm 0,2	19,7 \pm 0,4

^{*} $P < 0,001$; ^{**} $P < 0,01$; ^{***} $P < 0,01$

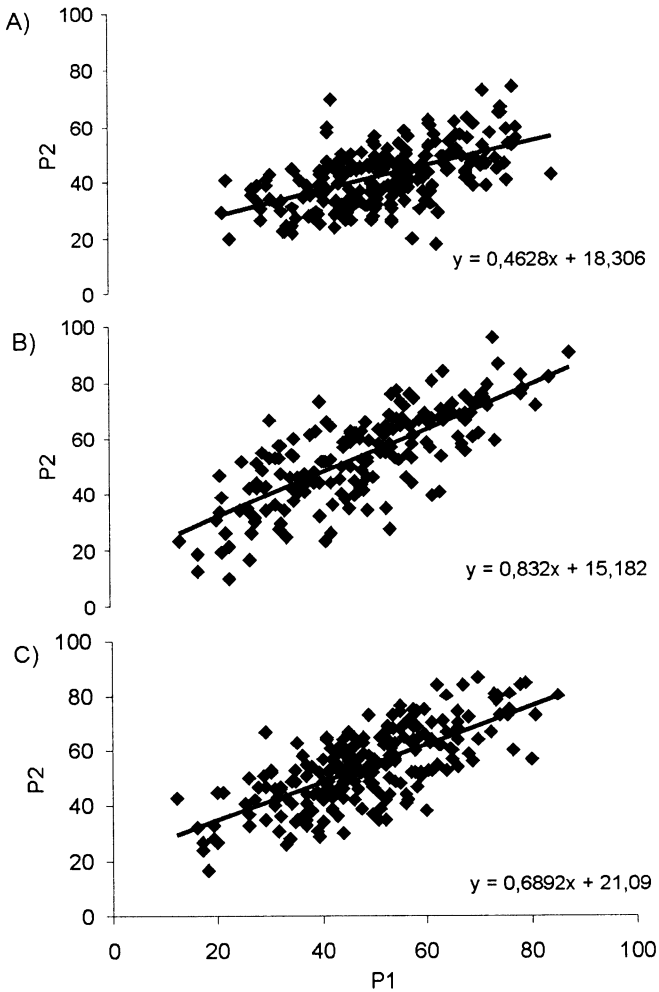


Figura 1. Resultados del segundo parcial teórico (P2) en función del resultado primero (P1) en los años 2001 (A), 2002 (B) y 2003 (C).

Cuando se compararon los resultados de cada parcial entre los 3 años se observó que el resultado del año 2002 fue significativamente diferente ($P < 0,0001$) para los dos parciales respecto a los resultados tanto del año 2001 como del 2003. No se observaron diferencias significativas entre el año 2001 y el 2003.

Cuando los resultados de la evaluación de los parciales teóricos se desglosan de acuerdo a la categoría se mantienen diferencias similares a las observadas para la globalidad (año 2001: Tabla 4, año 2002: Tabla 5, año 2003: Tabla 6).

Resultados del control práctico

El resultado de la primera y la segunda parte del CGP fue diferente en el 2002 ($P=1 \times 10^{-15}$), pero no fue diferente en el 2003 ($P=0,8$; Tabla 7). La pendiente de la relación entre los resultados de la primera y la segunda parte del CGP es significativa (2003, $P=8 \times 10^{-6}$, $r=0,29$; Figura 2A y B respectivamente).

Relación entre el resultado de los parciales teóricos y el control práctico

El resultado de la evaluación teórica fue significativamente mejor al finalizar el año que la evaluación práctica (2002: Tabla 8, $P=1,4 \times 10^{-30}$; 2003: Tabla 9, $P=4 \times 10^{-11}$). La curva de la relación entre el resultado de las evaluaciones teóricas y prácticas a lo largo del año es significativa en los 2 años analizados (2002: $P=9,3 \times 10^{-25}$, $r=0,60$; 2003: $P=6 \times 10^{-27}$, $r=0,64$; Figura 3A y 3B respectivamente).

Tabla 7. Puntaje promedio (\pm EE) en cada instancia de evaluación práctica (2002 y 2003).

	2002	2003
Primer parte CGP	51,4 \pm 1,3	45,2 \pm 1,4
Segunda parte CGP	63,1 \pm 1,3	44,9 \pm 1,3

Cuando los datos se descomponen en la primera y la segunda instancia de evaluación, observamos que diferencias entre el 2002 y el 2003. En el 2002 no existió diferencia significativa en la primera instancia, pero en la segunda los estudiantes obtuvieron un resultado significativamente mejor en la evaluación práctica (Tabla 8; $P=3,8 \times 10^{-51}$). En el año 2003 la diferencia existió en las dos instancias de evaluación, pero fue marcadamente mayor en la segunda (Tabla 9, primera instancia: $P=0,02$; segunda instancia: $P=9 \times 10^{-13}$).

Tabla 8. Porcentaje promedio (\pm EE) en cada instancia de evaluación y al finalizar el año, desglosado en la evaluación teórica y la evaluación práctica (2002).

	Primera instancia	Segunda instancia	Resultado anual
Parcial teórico	51,3 \pm 0,8	42,0 \pm 0,7	46,7 \pm 0,6
Parcial práctico	51,4 \pm 1,3	63,1 \pm 1,3	57,3 \pm 1,1

Tabla 9. Porcentaje promedio (\pm EE) en cada instancia de evaluación y al finalizar el año, desglosado en la evaluación teórica y la evaluación práctica (2003).

	Primera instancia	Segunda instancia	Resultado anual
Parcial teórico	47,8 \pm 1,0	54,9 \pm 0,1	51,0 \pm 0,9
Parcial práctico	45,2 \pm 1,4	44,9 \pm 1,3	45,0 \pm 1,3

Resultados de la evaluación por seminarios

En el año 2003, los resultados de los seminarios fueron significativamente mayores que los de los parciales (porcentaje: $90,6 \pm 0,6$ vs $48,6 \pm 0,9$, $P < 0,0001$). Los resultados de parciales y seminario y la diferencia entre los mismos fueron estadísticamente significativos para los estudiantes que finalizaron el curso en cada categoría ($P < 0,0001$; Tabla 10). La nota del seminario de las dos últimas categorías (Examen calificado y Exonera) fue significativamente mayor que la de Examen común y Pérdida de curso ($P < 0,05$), siendo las de estas dos ultimas categorías diferentes entre sí ($P < 0,05$).

Tabla 10. Porcentaje promedio (\pm EE) obtenido al finalizar el año en los parciales teóricos y en el seminario desglosado de acuerdo a la categoría en que el estudiante finalizó el curso (2003).

	Pérdida de curso	Examen común	Examen calificado	Exonera
Parciales	$27,8 \pm 1,2$	$46,8 \pm 0,5$	$61,0 \pm 0,6$	$75,2 \pm 0,9$
Seminario	$86,5 \pm 2,3$	$90,4 \pm 0,8$	$93,3 \pm 0,8$	$94,1 \pm 1,3$

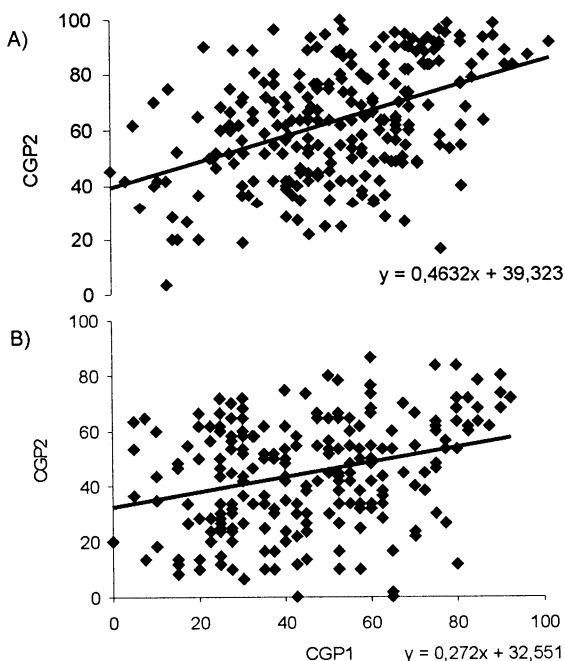


Figura 2. Resultado del segundo control globalizador práctico (CGP2) en función del resultado del primer (CGP1) en los años 2002 (A) y 2003 (B).

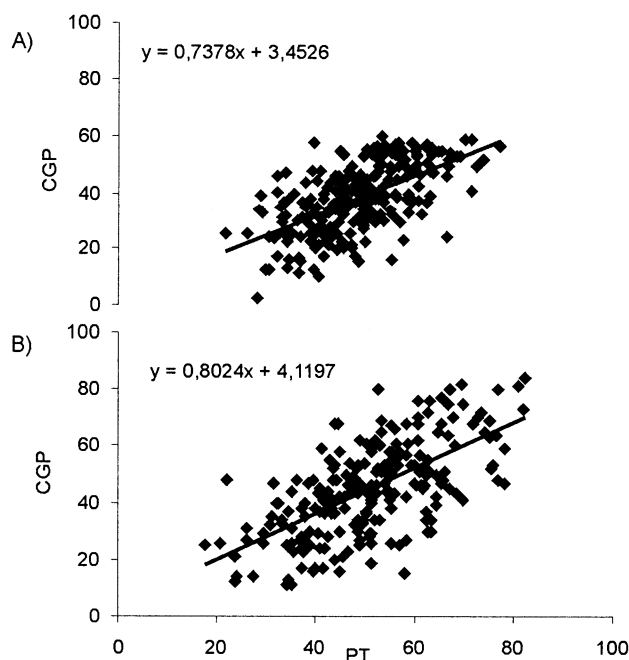


Figura 3. Resultado de la evaluación práctica (CGP) en función del resultado de la evaluación teórica (PT) en los años 2002 (A) y 2003 (B).

5. Discusión de los resultados

En primer lugar, es llamativa la diferencia en los resultados del rendimiento estudiantil de acuerdo al año. En el año 2002 se observaron resultados de parciales sensiblemente inferiores, además de diferir en la proporción de estudiantes que finalizaron el curso en cada categoría. A partir del análisis estadístico se puede concluir que las categorías Examen común y Examen calificado son las que generan la diferencia entre años. Sin embargo, dado que entre los 3 años analizados el 77,6% de los estudiantes se encontraron en dichas categorías es esperable que el análisis estadístico logre una mayor sensibilidad en la misma, no pudiendo a priori descartarse una incidencia de las otras dos categorías (15,5 y 6,9% para Pérdida de curso y Exoneran respectivamente).

El resultado académico de ambos parciales fue diferente, pero a su vez difirió para cada año. Mientras que en el año 2001 y 2003 el resultado mejoró en el segundo parcial teórico, en el año 2002 el resultado del primer parcial fue superior. Esto podría deberse a dos factores. Por un lado, la estructura del curso (orden de las Unidades Temáticas) fue diferente entre estos años, determinándose un mayor o menor peso

temático en uno u otro parcial. Por otro lado, se podría especular con que la diferencia podría deberse a las coincidencias entre fechas de parciales con otros cursos (no se cuenta con los datos). La tendencia se observa independientemente de la categoría en que hubieran finalizado los estudiantes, lo que sugiere que el resultado del primer parcial no se vinculó directamente con el del segundo. Esto a su vez es concordante con la correlación observada en el resultado del segundo respecto al del primero (ver Figura 1). Es llamativo que nuevamente en el año 2002 se observan diferencias, con un valor de r muy inferior al de los años 2001 y 2002. Esto indica que en dicho curso la predecibilidad del resultado del segundo parcial en función del resultado obtenido en el primero fue aún más baja que en los otros años. Esto sugiere que más que el reordenamiento del curso –que no tendría porque incidir en la correlación entre los resultados de ambos parciales– otros factores no considerados (como podrían ser las fechas de parciales y superposición con otras instancias de evaluación de otros cursos) determinaron el peor desempeño en el año 2002. Esta interpretación también es sostenida por las diferencias en los resultados para cada parcial de acuerdo al año, en que nuevamente el resultado del año 2002 difiere de los años 2001 y 2003 en ambos parciales.

El valor de r para la relación entre el resultado de la primera y la segunda parte del CGP es muy bajo, por lo que se puede afirmar que el resultado individual en la primera parte no permite especular con firmeza sobre el posible resultado del mismo estudiante en la segunda. A su vez, el mejor resultado observado en el segundo parcial teórico en el año 2003 no se observó en la segunda evaluación del CGP, lo que determina la importante diferencia observada entre el teórico y el práctico en la segunda instancia. En el año 2002 se observó un mejor resultado en la segunda instancia del CGP, mientras que fue el único año en que el resultado de la segunda evaluación teórica fue inferior al de la primera. Este resultado se contrapone a la posible interpretación anterior, de que factores externos como la fecha de parcial podrían haber afectado el resultado de la segunda instancia de evaluación en el 2002, sugiriendo que el problema estuvo centrado en la segunda evaluación teórica.

El resultado de los seminarios es sorprendentemente alto, más si lo comparamos con el de los parciales. Más aún, llama la atención lo pequeña de la diferencia del resultado obtenido en el seminario en función de la categoría en que finaliza el estudiante. Entre los estudiantes que perdieron el curso (no llegaron al 30% como promedio en los parciales) el resultado promedio del seminario es 86,5%, muy por sobre el porcentaje mínimo exigido para exonerar la materia. Es difícil encontrar una explicación única para un resultado tan llamativo, y la misma probablemente sea multifactorial (Ungerfeld, 2003). En primer lugar, la actividad de seminarios es sistemáticamente evaluada por los estudiantes como de las más motivadoras. La incorporación de estudiantes a actividades de investigación ya es valorada en forma importante como potenciador del aprendizaje en secundaria (Habraken et al., 2001),