



UNIVERSIDAD DE CHILE

Facultad de Ciencias Sociales

Departamento de Sociología

Sesgos en la evaluación docente: factores de género y ciclo de estudio

El caso de la facultad de ciencias sociales en la Universidad de Chile.

Tesis para optar al título de Sociólogo

Profesor Guía:

Rodrigo Asún Inostroza

Alumno Tesista:

Rodrigo Medel Sierralta

Santiago de Chile, Marzo de 2013

Agradecimientos

Excusando a la familia que siempre ha estado y estará conmigo cada día, me referiré únicamente a los dos pilares principales que hicieron posible que llevara a cabo este trabajo de tesis.

En primer lugar a mi profesor guía, Rodrigo Asún. Quién pese a la distancia y las dificultades que suponía dirigir un trabajo de tesis desde otro país, se comprometió y dedicó gran parte de su tiempo en guiarme por este proceso que duró casi dos años. No sólo tuvo la paciencia de enseñarme técnicas nuevas, respondiendo cada una de mis interrogantes, sino que además se daba el tiempo de discutir conmigo cada una de las dificultades, hallazgos y reflexiones que fueron surgiendo en el proceso investigativo. Le doy las gracias por creer en mis capacidades y por haberme convencido de que era capaz de llevar a cabo un estudio de gran nivel. Serán enseñanzas para toda la vida.

En segundo lugar a mi compañera de vida Rayen Cornejo Torres, por haber sido mi refugio emocional, mi desafío intelectual, mi libro abierto de sueños y aventuras. Gracias por haber estado a mi lado todo este tiempo dándome la confianza y el amor necesario para sacar adelante hasta los más complicados desafíos.

Índice

Capítulo 1 Planteamiento del problema	3
I. Formulación del problema.....	3
II. Objetivos de investigación	9
III. Relevancia del tema de investigación	10
Capítulo 2 Marco conceptual	11
I. Utilidad y validez de los cuestionarios	11
II. Factores extra clase, predisposiciones y/o sesgos	16
III. Ciclo de estudio y sesgo de género	23
Capítulo 3: Marco Metodológico	30
I. Hipótesis de investigación	30
II. Metodología de investigación.....	31
III. Plan de análisis	35
Capítulo 4. Resultados y análisis	42
I. Análisis Factorial	42
II. Análisis Univariado	46
III. Efectos de interacción	52
a. Análisis por ítem.....	52
b. Análisis agrupado por factor.....	
IV. Síntesis de resultados	83
Capítulo 5. Discusión y conclusiones	87
Bibliografía	96
Anexos	101

Capítulo 1. Planteamiento del problema

I. Formulación del problema

Las universidades son una de las instituciones más importantes y antiguas de las sociedades modernas. A lo largo de la historia han mostrado la flexibilidad y autorreflexión necesarias para ir adaptándose a los nuevos términos que imponen los cambios sociales, sin por eso perder sus características y objetivos principales (Tijssen & Weert, 2005).

De acuerdo al aún vigente análisis de las Universidades de Talcott Parsons y Gerald Platt, (citado en Tijssen & Weert, 2005: 124) las funciones de las universidades en términos típicos ideales se pueden dividir principalmente en cuatro:

- Producen conocimiento científico avanzado y entrenan a los jóvenes para adquirir esos conocimientos y convertirse en buenos investigadores.
- Proveen educación en general para formar a la ciudadanía
- Generan conocimiento sobre problemas de la sociedad humana a la vez que estimulan la capacidad crítica en los estudiantes para resolverlos.
- Preparan estudiantes para la práctica académica de sus profesiones, es decir, forman especialistas.

Mientras las primeras dos funciones provienen de la naturaleza primaria de las universidades, las últimas dos se derivan de los requerimientos adicionales que la sociedad le impone a las universidades.

Desde comienzos del siglo XIX ha habido una tensión y debate no resuelto sobre la importancia y predominio de cada una de estas cuatro funciones. Especialmente respecto a las cualidades intrínsecas de las universidades y la legitimidad y/o relevancia de las demandas externas (Tijssen & Weert, 2005).

Podemos entender que en el contexto de las funciones externas, es decir, la demanda social por el conocimiento especializado y tecnología para el desarrollo del país, la calidad de la docencia universitaria juega un papel estratégico al interior de las universidades. De ahí que se hayan ido incorporando distintos tipos de procedimientos para mejorar y asegurar la

calidad de la docencia dictada por las universidades, siendo uno de los más recurrentes la evaluación docente por parte de los alumnos.

La historia de la utilización de cuestionarios para la evaluación del desempeño del docente universitario tiene larga data. Se reconoce a Hernán Remmers como el creador del primer cuestionario de evaluación docente a fines de los años 20 en la Universidad de Purdue (Garduño, 2000). Desde ese momento, la evaluación docente ha sido uno de los temas más estudiados en el campo de la educación superior, con más de dos mil artículos al respecto hacia fines del siglo XX (García Garduño, 2000).

Sin embargo, no fue sino hasta la década de los setenta que dicha problemática se generalizó en la aplicación de Cuestionarios para la Evaluación Docente desde el Alumno (CEDA) en las universidades norteamericanas.

De acuerdo con Marsh (1984), los propósitos más habitualmente esgrimidos para justificar la aplicación de los CEDA, son los siguientes:

1. Diagnosticar y retroalimentar a los profesores en torno a su actividad docente, dando cuenta de su desempeño o efectividad docente.
2. Determinar méritos para la promoción y el otorgamiento de ciertos cargos de planta para los profesores.
3. Seleccionar los cursos e instructores por parte de los estudiantes.
4. Investigar los resultados y procesos docentes a partir de los resultados suscitados de estos cuestionarios.

A partir de la validez y confiabilidad que distintos autores le atribuyeron en un primer momento a estos cuestionarios (véase, Marsh 1984), se podría explicar su grado de generalización y difusión principalmente en Estados Unidos y Europa. Sin embargo, en gran parte del mundo, especialmente en Latinoamérica, aún existe mucha polémica respecto al grado de validez propia de esta evaluación; hay muchos mitos, temores y críticas que surgen tanto a nivel administrativo, en el profesorado y entre los mismos estudiantes. Este tipo de cuestionamientos nacidos en el seno de los estamentos

educacionales ha dificultado la aceptación de dicho tipo de cuestionarios en Chile, más no su aplicación.

De hecho, la reconocida utilidad de la evaluación docente a nivel internacional ha llevado a que a partir del año 2002 todas las universidades del consejo de rectores incorporen en su agenda institucional acciones que permiten desarrollar sistemáticamente dicha evaluación (Salazar, 2008). En efecto, el proceso de acreditación institucional –que surge a partir del año 2000, con la creación de la Comisión Nacional de Acreditación– puso como requisito la aplicación de esta evaluación al interior de las distintas casas de estudios. En este sentido, la evaluación por parte de los estudiantes se ha transformado en la forma más socorrida de evaluar la calidad del trabajo docente en el país.

Con respecto a los mitos que rodean la implementación de dichos cuestionarios, Aleamoni (1999) realizó un ensayo dónde refutó dieciséis de los más comunes. Concluye que estos mitos son, en general, efectivamente eso: mitos. El más común de todos es el que se relaciona con el hecho de que los estudiantes no pueden hacer juicios consistentes sobre el instructor y la instrucción que reciben debido a su inmadurez, la falta de la experiencia, y capricho (Almeaoni, 1999). Sin embargo, en una revisión bibliográfica realizada por el autor, este concluye que la literatura más reciente acerca del tema -Albanese (1991), Hativa (1996), y Palchik (1988)- indica lo opuesto, mostrando la estabilidad de las puntuaciones que los estudiantes asignan a sus profesores de un año a otro. En dicha puntuación se encuentran correlaciones en el rango de 0.87 a 0.89, lo que muestra una clara consistencia en las apreciaciones del alumnado.

Sin embargo, Almeaoni en este mismo estudio relaciona la disciplina académica del estudiante (factor que posiblemente es independiente de la calidad de la formación recibida), con los puntajes obtenidos en el test y efectivamente encontró diferencias en las puntuaciones asignadas por los estudiantes a los profesores de distintas disciplinas.

Factores como éstos son los que la literatura ha denominado “sesgos”, “predisposiciones” o “factores extra-clase” (Marsh 1984), refiriéndose a todas las variables que se encuentran más allá de la instrucción impartida y que afectan significativamente a la valoración que tienen los estudiantes de sus profesores. Si bien dichos factores son decisivos para estudiar

la evaluación docente, la investigación en torno a ellos ha sido bastante reducida en América Latina (García Garduño, 2000).

Lo anterior abre un abanico de preguntas sobre la pertinencia de incluir y considerar estos factores como influyentes o ponderadores a la hora de interpretar el sentido de las puntuaciones totales obtenidas por los profesores. En efecto, estos sesgos pueden significar que puntuaciones numéricamente distintas tengan igual significado en términos de calidad docente, en la medida en que exista una cultura evaluativa más laxa o más exigente en unas u otras disciplinas. De la misma manera, un mismo puntaje puede tener significados distintos si se toman en cuenta factores como el género del docente, la disciplina académica o ciertas características de los alumnos como el ciclo de estudios académico.

En el ámbito educacional chileno no existe registro de ninguna Universidad en la cual la evaluación docente incorpore, en el análisis de la calidad de sus profesores, factores como la facultad o disciplina en la que ellos se desenvuelven o características de los profesores y los alumnos que pueden influir en las evaluaciones, lo que podríamos concebir como un problema en la interpretación de los resultados de la evaluación de la docencia. Nos parece relevante para arrojar luz sobre la relevancia de esta carencia, entender cómo varían las percepciones de los estudiantes en función de variables que pueden estar asociadas a factores extra académicos.

Dada la amplia diversidad de factores extra clase que podrían afectar las evaluaciones de los profesores, nos parece interesante centrarnos en algunos factores en específico para poder analizarlos en profundidad. En específico, las características que nos interesan profundizar en este estudio son tres. Respecto a características del alumno nos interesa el género del alumno evaluador y el ciclo o nivel de estudios, mientras que respecto a características del profesor nos interesa el género del profesor. Para estudiar esto utilizaremos los registros obtenidos de la evaluación docente de la facultad de Ciencias Sociales de la Universidad de Chile del año 2007. Describiremos brevemente cada uno de estos factores para luego profundizar en cada uno de ellos en el marco teórico.

Nuestra variable central en el estudio será la de género, tanto del profesor como del alumno. Nos interesó esta variable en primer lugar por la relevancia que tiene el género

como variable explicativa de las conductas humanas y el creciente interés de la comunidad científica sobre este tema Pero además, el grueso de la investigación que estudia este punto concluye efectivamente que ambos sexos son evaluados de forma distinta (Garduño, 2000). Sin embargo, el que sean evaluados diferente no implica que uno de ellos sea siempre evaluado mejor que el otro. De hecho en un meta-análisis realizado por Feldman (1993), el autor encontró que las mujeres son mejor evaluadas en 28 de los 39 estudios analizados. Resultados que variaron según los distintos contextos institucionales.

Siguiendo esta línea, la socióloga Estadounidense Susan Basow (1995) confirmó los hallazgos de Feldman respecto a la correlación entre los puntajes y el género del docente señalando que si bien no muestran tendencias claras hacia uno u otro lado, los investigadores han cometido el error de observar siempre los puntajes totales. Su argumento señala que es en la interacción entre preguntas individuales con otras variables externas donde se pueden observar patrones de género. Por lo tanto la investigadora analizó los puntajes totales en función de ciertas variables de segmentación como las distintas disciplinas que cursan los estudiantes y su relación con los docentes y su género. Ahí se encontró con que los profesores hombres en ciencias sociales eran mejor evaluados que las profesoras en dimensiones específicas como: entusiasmo, estimulación y retroalimentación. También observó diferencias en los puntajes obtenidos en Ciencias donde la evaluación general de los hombres es más alta que el de las mujeres. Este estudio inicial fue replicado por la autora en diversas Universidades de Canada y EEUU (Basow 1995; Basow 2000; Basow 2006) encontrando siempre que existe una interacción significativa entre el género del profesor y el del estudiante. En general, las profesoras mujeres tendían a recibir sus calificaciones más altas de estudiantes mujeres y sus calificaciones más bajas de los estudiantes varones. Seguiremos en ese sentido la recomendación de Basow y buscaremos desentrañar la influencia de la variable de género del profesor en sus interacciones, principalmente con una que llamamos “ciclo de estudios” y con el género de los alumnos.

Respecto al factor correspondiente al año cursado o ciclo de estudios, el cuál ha sido el factor extra clase que más ha demostrado influir en la evaluación docente (Hills, Naegle y Bartkus 1992), lo tomaremos como variable contextual, toda vez que nos permitirá ver cómo se comporta el sesgo de género en sus distintas categorías o ciclos de estudio. Pero

también la tomaremos como variable individual, es decir, para ver el efecto que produce por sí sola. En ese sentido los investigadores han demostrado que los estudiantes de primer o segundo año tienden a estar más preocupados de asuntos de la estructura del curso, tales como existencia de una carga de trabajo adecuada y la pertinencia de materiales y exámenes, mientras que los estudiantes de los últimos años, usualmente dan una mayor importancia a la discusión en clases, elemento que según los autores coincide con la mayor participación y naturaleza comprometida de muchos cursos superiores (Hills, Naegle y Bartkus, 1997).

Las características de las variables incluidas en el cuestionario de evaluación docente de la Facultad de Ciencias Sociales el año 2007, nos permiten explorar una interacción de la que no se tiene registro aún en la literatura Latinoamericana sobre el tema, que es justamente la relación entre el género (del profesor y del alumno) y el nivel o año cursado por el alumno.

Hay que tomar en cuenta que el gran cuerpo de estudios sobre sesgos de género en la evaluación docente ha encontrado que, si bien las diferencias entre ambos géneros existen, el tamaño de sus efectos es pequeño, como es común en el ámbito de la educación superior (Basow, 1998). Por lo tanto, más que esperar grandes tamaños de efectos estadísticos, nos interesa sobre todo que los hallazgos sean significativos y consistentes entre sí.

II. Objetivos de investigación

Objetivo General:

Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile se encuentran asociados al género del alumno y del profesor y determinar en qué medida estos sesgos varían a lo largo de los distintos niveles de estudio.

Objetivos específicos

1. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile se ven afectados por el género del profesor.
2. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile se ven afectados por el género del alumno.
3. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile se ven afectados por el ciclo de estudios del alumno.
4. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, se ven afectados por la interacción entre el género del profesor y el género del alumno.
5. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, se ven afectados por la interacción entre el género del profesor y el ciclo de estudios del alumno.
6. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, se ven afectados por la interacción entre el ciclo de estudios y género del alumno.
7. Conocer en qué medida los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, se ve afectada por la triple interacción entre el género del profesor, del alumno y el ciclo de estudios cursado por el alumno.

III. Relevancia del tema de investigación

a. Relevancia teórica

La revisión bibliográfica afirma tanto la validez de la evaluación docente como su utilidad. Sin embargo, existe ya una considerable investigación empírica que ha cuestionado esta validez a la vez que demostrado que existen variables que inciden en la evaluación que los estudiantes realizan de sus profesores. La investigación ha puesto especial atención a características del instructor, del curso o disciplina académica, y del alumno. Estas características influyen, en mayor o menor medida, determinadas por el entorno, en las calificaciones que el estudiante realiza de la instrucción recibida. A estas características se les suele llamar factores extra clase o sesgos.

Existe evidencia, entre la que cabe mencionar el estudio realizado por los investigadores Joey Sprague y Kelley Massoni (2005), que sugiere la existencia de un sesgo de género encubierto en las evaluaciones docentes. Por tanto, vale la pena tomar en cuenta factores de sesgo asociados tanto al género de los profesores como de los alumnos y ver la interacción entre ambos.

También vale la pena poner especial énfasis en características asociadas a los alumnos, especialmente las que dicen relación con el año de estudios cursado. Los Canadienses Smith y Cranton (1992) realizaron un estudio que examina las percepciones de los estudiantes en diferentes dimensiones de la enseñanza según los distintos niveles de instrucción o año de estudios cursado por los alumnos. En este estudio demostraron que el puntaje obtenido por los profesores en relación a distintas dimensiones de la efectividad docente varía según el año de estudios que se cursa. Faltaría ver, por tanto, en qué dimensiones pueden darse estas diferencias y si pueden o no estar relacionadas con los sesgos de género que estudiaremos, para lo cual se cuenta con un acercamiento por medio de datos cuantitativos correspondiente a los puntajes obtenidos por los profesores en la evaluación docente.

En este sentido nos proponemos innovar al buscar relacionar dos ámbitos, que serán el analizar posibles asociaciones y diferencias con sesgos asociados a características tanto del

alumno como del profesor en función del género de ambos, a la vez que buscaremos explorar si modula esta asociación el año que cursa el estudiante. Nos parece necesario elaborar un estudio de esta clase ya que hasta la fecha no se ha realizado una investigación sobre la influencia de este tipo de factores extra-clase en la evaluación de los docentes en el país, pese a la creciente relevancia de los estudios de género en las ciencias sociales Latinoamericanas.

2.2 Relevancia metodológica y práctica:

La relevancia metodológica que nos parece más destacable en la presente investigación se refiere a la posibilidad de utilizar, en el transcurso de la investigación, análisis avanzados de estadística multivariable y métodos cuantitativos poco utilizados en Chile como es el caso de construcción de escalas utilizando procedimientos factoriales para datos ordinales, y en segundo lugar realizar análisis de varianza univariados y multivariados para relacionar las variables independientes y dependientes.

Por otra parte, a nivel práctico, nos parece fundamental el desarrollo de esta investigación ya que se trata de una temática poco estudiada, no solo en Chile dónde un estudio de este tipo es inédito, sino también en América Latina, contexto en el que las evaluaciones docentes por parte de los estudiantes son cada vez más relevantes. Esto, para el caso de la Universidad de Chile, constituiría una buena base tanto para futuras investigaciones como recomendaciones y ajustes en los instrumentos de evaluación docente, en función de los resultados y conclusiones de este estudio.

Capítulo 2: Marco conceptual

I. Utilidad y validez de los cuestionarios de evaluación docente.

A pesar de que muchas universidades han adoptado la evaluación docente como una (y menudo la más influyente) medida para controlar la calidad de la instrucción, últimamente ha aparecido una interesante polémica sobre la validez de estos cuestionarios. Si bien se ha demostrado que estos cuestionarios tienden a entregar una medición válida de la calidad de la docencia y que permiten proveer insumos relevantes para el mejoramiento del desempeño de los profesores (García Garduño, 2000), la investigación reciente, principalmente la de los últimos 10 años, establece algunas matizaciones sobre estas aseveraciones (García Garduño, 2008).

Respecto a los estudios que afirman la validez de los cuestionarios, hay autores que han demostrado por medio de distintos estudios que estos métodos de evaluación son válidos a nivel global. Uno de los tantos procedimientos utilizados para analizar su validez e impacto han sido los estudios de retroalimentación de las evaluaciones (Marsh y Roche 1997). En este tipo de estudio los instructores se asignan de manera aleatoria tanto para participar en grupos de tipo experimental (retroalimentación) como en grupos de control.

Profundizando lo anterior, el procedimiento que se sigue es el siguiente: una vez que el cuestionario ha sido respondido, se procede a devolver lo más rápidamente los SETS (Student Evaluation of Teaching) a los instructores del grupo experimental y se observa el efecto que éste insumo tiene en su futuro desempeño. El meta-análisis realizado por Cohen (1990), comprobó que la evaluación que se hacía a los instructores en los grupos de retroalimentación presentaba desviaciones estándares porcentualmente más altas que los controles (a quienes no se devolvían los resultados). Por otra parte, las diferencias más significativas eran observadas en ciertos factores que terminaban siendo decisivos, tales como: las habilidades del instructor, la actitud hacia la materia que presentaban los estudiantes, y la retroalimentación a los estudiantes.

Efectivamente, en las contribuciones del estudio de Furhmann y Geyer (2002), se subraya la utilidad e importancia de las evaluaciones de los estudiantes hacia la instrucción recibida.

Sus consideraciones prosiguen arguyendo que el puntaje global obtenido por los maestros depende estrechamente de la percepción de los estudiantes sobre el desempeño del profesor en la clase, es decir de la percepción del alumnado en torno a la instrucción recibida. Principalmente aspectos como: “aclarar las explicaciones”, “ejemplos concretos” y “retroalimentación en su progreso de aprendizaje”, resultan fundamentales para asegurar un buen desempeño y por ende un buen puntaje por parte del docente. Por otro lado, criterios como “amistad” o “paciencia” resultan menos relevantes. En este punto, se aprecia la importancia de dicha evaluación para dar cuenta de la realidad y apreciación del proceso de aprendizaje por parte de los estudiantes. En este sentido, los maestros han comenzado a utilizar dicha evaluación del alumnado como retroalimentación que permite deslindar de qué manera ha sido traspasado el contenido del aprendizaje desde el punto de vista de los estudiantes (Furhmann y Geyer, 2002).

Otro de los más importantes estudios que busca demostrar la confiabilidad y validez de los instrumentos de evaluación docente, es el realizado por Marsh y Roche (1997). En esta investigación realizan una revisión de artículos que afirma que en general bajo ciertas condiciones apropiadas, las evaluaciones de estudiantes de la enseñanza son: (a) multidimensionales; (b) confiables y estables; (c) primariamente una función del instructor que enseña un curso antes que del curso que es impartido; (d) relativamente válidas respecto de una serie de indicadores de la enseñanza efectiva; (e) relativamente inafectadas por una variedad de variables hipotéticas como las predisposiciones potenciales (por ejemplo, tamaño de clase, carga de trabajo, interés previo por la materia); y (f) útiles en cuanto a la capacidad de mejorar la eficacia de la docencia cuando los SETS se asocian a un sistema de consultorías y asistencias al docente.

Sin embargo, como ya hemos dicho, recientemente han ido apareciendo nuevos estudios, que si bien no invalidan los resultados obtenidos por Marsh, ni la pertinencia de los cuestionarios, permiten observar ciertos sesgos y factores que afectan a las evaluaciones y que van más allá de la instrucción recibida por el estudiante. Este tema resulta decisivo cuando pretendemos explicar y dar cuenta las diferencias que se pueden apreciar en los puntajes obtenidos por distintos profesores. En efecto, si bien Apollonia y Abrami (1997) coinciden con las apreciaciones generales de Marsh, al presentar la evidencia de que la

instrucción es multidimensional, muestran los resultados de un meta-análisis sobre la validez, el cual indica que las clasificaciones del estudiante son sólo moderadamente válidas. Los autores observan que ciertos factores como las características del instructor (grado académico), y características del curso (tamaño y la sección) influyen en alguna medida en las apreciaciones que el estudiante realiza sobre la instrucción recibida.

En un estudio reciente, García Garduño (2008) señala que uno de los problemas de fondo de los cuestionarios de evaluación es su indeterminación. Es decir no presenta una explicación unívoca de un cuerpo de datos. Esto ocurriría cuando el instrumento presenta sólo reactivos generales cuyos estándares mínimos son fijados arbitrariamente por la universidad sin existir un concepto claro de lo que es la eficacia o calidad docente.

En este contexto cobra especial trascendencia el estudio realizado por Onwuegbuzie y colaboradores (2007), el cual nos muestra que los más graves problemas de los cuestionarios de evaluación son su validez de contenido (es decir, validez de los ítems) y de constructo (es decir, la validez de fondo estructural, de resultados, de generalización).

Este estudio utilizó un análisis multietápico con métodos mixtos para evaluar la validez de contenido y de constructo relacionado de un formulario de evaluación de la docencia de una universidad pública, esto mediante el examen de las percepciones de los estudiantes sobre las características de los profesores universitarios. Los participantes fueron 912 estudiantes de pregrado y de posgrado (10,7% del total de estudiantes) de diversas especializaciones académicas. Onwuegbuzie y colaboradores (2007) señalan que si bien los cuestionarios pueden tener validez aparente, la validez real del instrumento no siempre es igual a lo aparente, en la medida en que sus ítems pueden parecer relevantes en mayor o menor medida dependiendo de quién lo contesta. En otras palabras, no existen suficientes evidencias de que los ítems de los cuestionarios tengan suficiente validez de contenido, ya que no tienen definido con suficiente claridad el concepto de efectividad docente. De hecho el análisis que ellos realizaron mostró que tres de los temas más frecuentemente mencionados por los alumnos de esa universidad como constituyentes de un buen docente, no estaban representados por ninguno de los elementos del cuestionario de evaluación que allí se aplicaba.

Parece que hay una clara diferencia entre lo que los desarrolladores del cuestionario de evaluación consideran las características de los instructores efectivos y lo que los estudiantes consideran las características más importantes. Los autores concluyen que esta brecha sugiere que los criterios de evaluación de estudiantes hacia sus profesores universitarios no pueden ser adecuadamente representados por los cuestionarios de evaluación, lo que podría afectar negativamente a la capacidad de los estudiantes de realizar una crítica integral de sus profesores. Así, aunque los puntajes obtenidos por medio de esta forma de evaluación universitaria son fiables, la validez del puntaje general puede ponerse en cuestión. En una época en que se utiliza la información obtenida por estos cuestionarios para tomar decisiones sobre la tenencia, la promoción, y las cuestiones de pago por mérito, esta amenaza potencial para la validez es inquietante y merece más investigación.

Respecto a la validez de constructo, Onwuegbuzie (2007) señala que la mayoría de los temas evaluados varían en relación a los atributos del estudiante (por ejemplo, el género, la edad), poniendo en tela de juicio la validez de constructo de las puntuaciones de los cuestionarios. Las relaciones encontradas entre la mayoría de las variables demográficas y varios de los temas y meta-temas, sugieren que a la hora de interpretar las respuestas y los puntajes obtenidos, los administradores deben considerar el perfil demográfico de la clase subyacente. Desafortunadamente, esto no parece ser la práctica actual. Por ejemplo, la constatación de que las estudiantes mujeres tienden a poner más peso sobre factores como la atención que ponen los profesores a los estudiantes, sugiere que una clase con estudiantes de sexo predominantemente femenino podría dar mayor importancia a estos ítems que una clase que contiene sobre todo estudiantes varones.

En síntesis podemos señalar que se ha demostrado que los cuestionarios de evaluación docente son una buena herramienta de medición con buen impacto en el rendimiento y en la calidad docente. Sin embargo, la investigación reciente también indica que es necesario evaluar si están sesgados o no antes de usarlos ya que de lo contrario se puede obtener información poco imparcial que pone en riesgo su validez.

En otras palabras, hay que estudiar a fondo la validez de contenido por medio de una revisión de los conceptos de efectividad docente y la validez de constructo por medio de una revisión de los factores extra clase o sesgos. Nosotros nos enfocaremos en el segundo

punto, es decir en los factores que pudiesen afectar en mayor o menor medida la validez del constructo, para lo cual centraremos la atención en las características del alumno, del profesor y del contexto en el cual estos cuestionarios son aplicados.

A continuación iremos ahondando este tema de manera de ir esclareciendo y problematizando dicha temática.

II. Factores extra clase, predisposiciones y/o sesgos

Cuando nos referimos a los factores extra-clase estamos refiriéndonos a todas las variables que se encuentran más allá de la instrucción impartida y que afectan significativamente la evaluación docente. A dichas variables se les suele denominar: sesgos, predisposiciones o factores extra-clase (Marsh 1984). Si bien dichos factores son decisivos para estudiar la evaluación docente, la investigación en torno a ellos, ha sido bastante reducida en América Latina (García Garduño, 2000).

En este punto Vásquez y Gabalán (2006), se interrogan acerca del origen de estos sesgos. A partir de esta interrogante, afirmarán que el análisis de las interrelaciones entre profesores, estudiantes, institución y entorno social, constituye un buen punto de partida para determinar qué aspectos, según los estudiantes, son los más relevantes al momento de emitir juicios evaluativos sobre el desempeño general de sus docentes, ya que dichos factores afectan, como ya se ha señalado, de una u otra forma su percepción sobre la instrucción recibida. En este sentido, podemos concluir que la evaluación docente es un producto de la interacción de los factores educativos tales como: el profesor, el estudiante y la institución, los cuales en un contexto determinado afectarán en mayor o menor medida la percepción del alumno al momento de evaluar el desempeño de su profesor, lo que daría origen a los sesgos o factores extra-clase (Vásquez y Gabalán, 2006).

A pesar de que se ha estudiado y demostrado la existencia de estos factores o sesgos, la investigación tradicional sobre el tema señala que estos factores no son lo suficientemente fuertes como para invalidar un instrumento de evaluación docente. Los primeros estudios importantes realizados sobre esta materia han llegado al consenso de que los factores extra-clase juegan un papel menor en las clasificaciones del estudiante sobre su propia instrucción (Marsh, 1984)

Sin embargo con el tiempo, principalmente los estudios de la última década (véase Onwuegbuzie, 2007; Basow 2000) han ido dejando al descubierto lo significativo que resultan estos factores en las puntuaciones finales que los estudiantes hacen de sus profesores. Dichas variables, que influyen en mayor o menor medida en la evaluación docente, se pueden agrupar en tres tipos: las características del profesor, las características del alumno y finalmente las características de los cursos y de las disciplinas (Garduño 2000). Cada uno de estos factores generaría ciertos sesgos o predisposiciones que los investigadores han ido observando y estudiando a lo largo de los años.

A continuación desarrollaremos brevemente en que consiste cada uno de estos grupos tomando en cuenta los resultados de las principales investigaciones sobre el área, dejando para el final los sesgos de género y el nivel de estudio que constituyen los factores centrales de la presente investigación.

A. Características de los cursos y las disciplinas.

Hallazgos en distintas investigaciones demuestran que efectivamente existen diferencias en la evaluación promedio de los profesores según características de los cursos y las disciplinas. Las características más influyentes serían:

Tamaño del grupo: En un meta análisis, Feldam (1984) encontró que la relación entre tamaño del curso y la evaluación es inversa, es decir, los grupos de menor tamaño tienen una mejor evaluación que los de mayor tamaño. Además, encontró una relación en forma de U, es decir los de mayor tamaño también evalúan mejor que los de tamaño medio. En este sentido, Marsh y colaboradores (1979) ya había hallado algo semejante, pero con la especificidad de que los cursos de menor tamaño califican mejor en la interacción con el profesor que los grupos grandes.

Naturaleza de la disciplina: La investigación ha demostrado en diversos estudios que la evaluación global y por dimensiones puede variar según la disciplina.

En general, las disciplinas científicas son peor evaluadas que las humanistas y de ciencias sociales (Garduño, 2000). Lo anterior quiere decir, por ejemplo, que un profesor de ingeniería con promedio 3,3 no es necesariamente inferior que uno de filosofía con

promedio 4,3. Feldman (1978), muy tempranamente llegó a estas mismas conclusiones, es decir, que las carreras humanísticas reciben mejores puntuaciones promedio si se comparan directamente con los puntajes obtenidos por las carreras científicas.

Continuando con esta línea, los autores Neumann y Neumann (1983) realizaron un estudio de diferencias disciplinarias basadas en el modelo de “Biglan”, que, se inspira en algunas ideas de Khun para clasificar las disciplinas según sean paradigmáticas (duras) o pre-paradigmáticas (suaves). La conclusión final del estudio fue que no puede existir un instrumento único de evaluación para las distintas disciplinas. En las disciplinas suaves o pre-paradigmáticas, las habilidades docentes del profesor son más importantes que en las paradigmáticas debido a que en la primera hay menos desarrollo de teoría o paradigmas. Es decir, la habilidad docente es más importante en las humanidades que en las disciplinas científicas.

En el estudio de Hills, Naegle y Bartkus, (1997), los autores, luego de examinar a 484 estudiantes de una gran universidad pública especializada principalmente en economía, realizaron una comparación de las diferencias por programas académicos según su especialización. Los autores crearon tres grupos: los estudiantes de negocios, estudiantes materias distintas a los negocios, y los no reportados o estudiantes no declarados. Los autores encontraron importantes diferencias entre los estudiantes que eran de negocios y los estudiantes que no eran de negocios, particularmente siete ítems (de los 19) fueron declarados por los segundos y no por los primeros como los más importantes en el proceso de enseñanza. Entre esos estaban por ejemplo: la “organización de la clase”, “el instructor da la oportunidad para expresar opiniones y comentarios en la clase”, y “el instructor ayuda a resolver dudas”. Esto sugiere que los estudiantes de negocios son inherentemente diferentes de sus colegas de otras disciplinas y pueden ser menos críticos de los entornos ambiguos o de la poca participación en clase.

Además, los estudiantes que tienen todavía que declarar una especialidad, encuentran que la claridad de los profesores para señalar las responsabilidades del estudiante es de mucha más importancia que lo que piensan los estudiantes de negocios. Estos resultados proporcionan ciertas sugerencias interesantes en cuanto a cómo estudiantes perciben y evalúan los ítems en un cuestionario según su especialización o disciplina.

Por su parte, la investigación realizada por Rizo y Coello, (2006), en la Universidad autónoma de Occidente en Colombia, analizó a nivel cuantitativo el grado de relevancia de las distintas dimensiones medidas por los instrumentos de evaluación docentes, según los distintos programas académicos. Los autores lograron establecer conclusiones particulares tales como que en la Facultad de Ciencias Económicas y Administrativas, los factores más relevantes son: “el profesor facilita el aprendizaje de los estudiantes” y “las calificaciones asignadas por el profesor han sido justas”. Mientras que en otros programas como los de la Facultad de Comunicación Social la variable más relevante es “el profesor manifiesta en su discurso y en la manera de llevar el curso, dominio sobre el campo de conocimiento de su asignatura”.

Estos son elementos importantes a tener en cuenta dado que el estudio se realizará en la facultad de Ciencias Sociales de la Universidad de Chile, por lo que se puede suponer que los resultados no son necesariamente extrapolables al resto de la comunidad universitaria.

B. Características del profesor

Respecto a los sesgos que surgen de las características del profesor, la investigación se ha centrado en ciertos factores tales como: la personalidad del profesor, la relación entre la productividad de su investigación y su docencia, el grado académico obtenido, su edad, experiencia, las calificaciones que otorga a sus alumnos y su género. Éste último punto lo trataremos en detalle más adelante. A continuación realizaremos una revisión de cada uno de los factores anteriores:

- **La personalidad del profesor:** Recién a partir de los años 70 y 80 la investigación comienza a interesarse en factores como la personalidad del profesor al observar la alta correlación que se podía establecer entre ésta y la efectividad de su docencia. García Garduño (2000) realiza una síntesis de los principales hallazgos encontrados hasta la fecha en cuanto al tema, concluyendo que los rasgos fundamentales que inciden, tomando en cuenta la personalidad de un docente, en la eficacia de su docencia son: liderazgo, extroversión, solidaridad, inteligencia y curiosidad intelectual.

- **La relación entre la productividad de la investigación y la docencia.** Siguiendo a Garduño (2000), podemos afirmar que los estudios realizados recientemente han

confirmado una relación positiva entre investigación y efectividad docente. Respecto a esto Feldman (1987) realizó un meta análisis donde analizó 43 investigaciones publicadas entre 1950 y 1983, encontrando un coeficiente de correlación pequeño pero significativo entre productividad académica y efectividad docente. Dicha investigación sumó un nuevo elemento que resultó significativo al incorporar la variable “disciplinas académicas”. En este punto, se pudo concluir que la productividad en investigación facilitaba la efectividad docente en las Ciencias Sociales y Humanidades pero no así en las Ciencias Naturales.

- **Otros factores:** Respecto a otros elementos que pudieran estar asociados a las características del profesor, Garduño (2000) afirma que si bien existen otros factores que inciden en esta temática, serían menores. Entre estos habría que considerar: el grado académico, la edad, la experiencia y las calificaciones que los profesores asignan a sus estudiantes. La investigación más completa, a nuestro parecer, en torno a estos factores es el meta-análisis de Feldman (1983), el cual confirmó que las relaciones no eran muy significativas en cada uno de los puntos estudiados. Sin embargo, aún si los coeficientes de correlación son bajos, se puede afirmar que efectivamente los profesores más “blandos” o indulgentes, es decir, los docentes que suelen evaluar mejor a sus estudiantes, son a su vez mejor evaluados por éstos, aunque la relación es débil y no siempre es lineal.

c. **Características del alumno**

Hay un relativo consenso respecto al hecho de que las percepciones de los estudiantes en el momento de evaluar a un profesor, más allá de las predisposiciones potenciales de los estudiantes, se ven determinadas principalmente por el desempeño del docente y la capacidad del profesor de entregar conocimientos relacionados con su disciplina (Furhmann y Geyer, 2002). En el estudio de Furhmann y Geyer, (2002), los autores analizan el efecto de las predisposiciones potenciales en los puntajes globales asignados a los profesores en el contexto de la enseñanza. La investigación toma en cuenta predisposiciones como el interés de los estudiantes en la materia estudiada y dirán que éstos pueden ser un resultado del buen desempeño en la enseñanza impartida y no necesariamente una predisposición simple del estudiante. Estos autores realizan una investigación basada en 40 entrevistas cualitativas con estudiantes de las universidades comerciales austríacas. Los resultados muestran que

una buena evaluación de los estudiantes hacia sus maestros, dependerá principalmente de la calidad de la instrucción impartida. Por su parte, los modelos de ecuaciones estructurales muestran que los puntajes globales de los maestros dependen principalmente de su comportamiento en el proceso de enseñanza (Furhmann y Geyer, 2002).

Pese a este consenso respecto a que el factor más relevante va a ser siempre la calidad de la instrucción recibida, las clasificaciones globales se ven afectadas también por las actitudes de los estudiantes en relación a la nota obtenida, su interés previo en la disciplina, al año o nivel académico y su género.

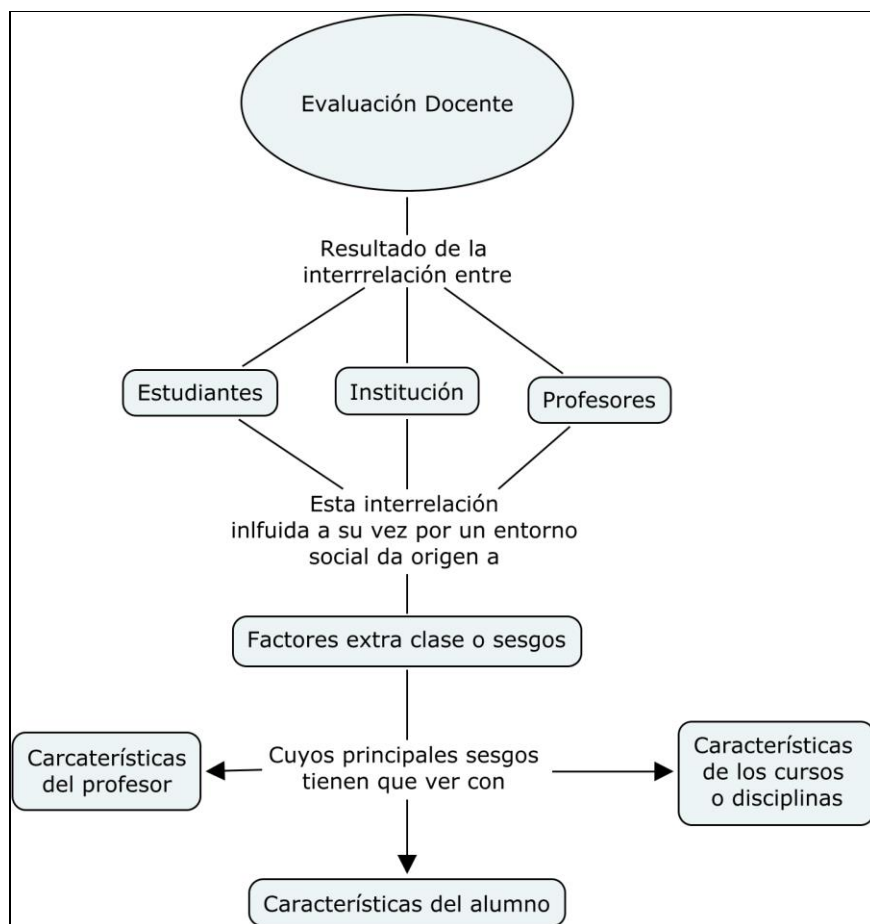
En el estudio de Hills, Naegle y Bartkus, (1997), los autores sugieren que la importancia de un ítem para el estudiante, depende de las características propias del estudiante. Luego de examinar a 484 estudiantes de una universidad pública, sugieren que no todos los ítems son mirados de igual forma y que medidas como la relevancia del ítem puede diferir según las características del estudiante tales como género, año de estudios, y notas obtenidas. Respecto a las notas, los autores dirán que estos juegan un papel de mediana importancia en el proceso. Son los estudiantes de la parte superior (las mejores notas) los que dan mayor importancia a ítems tales como “la claridad de expectativas” y “disponibilidad de la ayuda”. Los autores señalan que tal vez la habilidad para hacer uso de un instructor y aclarar los problemas y su resolución, es lo que permite a los estudiantes mejores mantener sus calificaciones altas, lo que explicaría sus prioridades en materia de ítems (Hills, Naegle y Bartkus, 1997).

Por otro lado, los resultados del estudio de Barktus y colaboradores (1997) sugieren que en conjunto estudiantes valoran la “transparencia en sus cursos”. Les gusta saber cómo están siendo evaluados, que esa evaluación sea justa y que el trabajo asignado sea apropiado para las metas del curso.

Por último, el interés previo por tomar el curso es uno de los factores que más impacta la evaluación docente. Este factor explica el doble que la nota esperada por el alumno. Los hallazgos de Marsh (1982) señalan que el interés por tomar el curso es lo que explica que a mayor carga de trabajo, mejor sea la evaluación obtenida por el profesor.

En los apartados anteriores se ha hecho un esfuerzo por sistematizar y ordenar la evidencia más relevante respecto de los factores extra clase (ver Figura 1). A continuación profundizaremos en los factores centrales para el estudio, los cuales tienen relación tanto con características propias del profesor como del alumno.

Figura 1: Factores extra clase



Fuente: Elaboración propia

III. Ciclo de estudio y sesgo de género

A. Ciclo de estudios

El factor perteneciente a las características del alumno que más ha demostrado influir en la evaluación docente, dice relación con el año cursado o ciclo de estudios de la carrera. En el estudio ya citado de Hills, Naegle y Bartkus (1992), los autores demostraron que los estudiantes del primero o segundo año estaban más interesados con los asuntos de la estructura del curso, tales como existencia de una carga de trabajo adecuada y la pertinencia de materiales y exámenes, mientras que los estudiantes de los últimos años, por otro lado, dieron una mayor importancia a la discusión, elemento que según los autores coincide con la mayor participación y naturaleza comprometida de muchos cursos superiores (Hills, Naegle y Bartkus, 1997, 2009).

En esta línea, los investigadores Canadienses Smith y Cranton (1992) realizaron un estudio que examina las percepciones de los estudiantes en diferentes dimensiones de la enseñanza según los distintos niveles de instrucción o año de estudios cursado por los alumnos. De esta manera, se reunieron puntajes de estudiantes de 2816 cursos en una facultad de comercio Canadiense de tres semestres distintos de pregrado y otro semestre de posgrado. En la muestra combinada, los autores encontraron diferencias en las percepciones de los estudiantes para todos los factores según el semestre cursado. Es decir, el puntaje obtenido por los profesores en relación a distintas dimensiones de la efectividad docente varía según el año de estudios que se cursa. Los autores afirman que en los primeros años, factores como organización y claridad son mejores predictores de la puntuación del docente que en los años posteriores, donde destacan factores como atmósfera y evaluación.

En el presente estudio, junto con adentrarnos a explorar si efectivamente las puntuaciones generales de los profesores varían a lo largo de la vida universitaria de los alumnos, también analizaremos las interacciones que se pueden dar entre este ciclo de estudios con las otras dos variables independientes que serán el “género del alumno” y el “género del profesor”. A continuación profundizaremos las variables relacionadas al género, tanto del profesor como del alumno.

B. El sesgo de género

b.1. Sobre diferencias de género en la educación superior

Antes de referirnos a la relación específica entre género y evaluación docente hay que primero mencionar algunos aspectos más generales sobre los estudios que se han hecho sobre las relaciones de género en el ambiente académico universitario.

En primer lugar hay que señalar que si bien las trabas formales para la participación tanto de académicas como de estudiantes mujeres en la educación superior ya no existen, no por ello el espacio de la educación superior se transforma en un lugar neutro de predisposiciones de género. Sin bien la educación superior es un espacio más reflexivo que permite un mayor intento de control de estos mecanismos y representa una oportunidad para adquirir conocimiento no transmitidos por la familia o grupos de pares, ello no implica que estas diferencias no se den en la práctica (Lamadrid, 2012). Las instituciones de educación, como han señalado Bourdieu y Passeron (1995), son en sí mismas mecanismos de reproducción cultural, social y valórica. Se refuerzan valores dominantes lo que lleva inevitablemente a transmitir patrones de género existentes en la sociedad.

Sobre nuestro contexto local, hay que partir del hecho de que el ingreso de mujeres al ámbito académico Chileno ha sido relativamente reciente. Ejemplo de ello es que entre los académicos menores de 35 años las mujeres representan el 47%, mientras que en el grupo de 65 o más años sólo el 22% (Olavarría, 2012). Respecto a los cargos directivos Saracosti (2006) señala que las diferencias son muy abismantes, tanto así para que el 2005 las mujeres representan sólo el 8% del total de rectores de las Universidades en Chile.

Dentro de los pocos estudios que hay en Chile sobre diferencias de género en el mundo académico universitario, está el estudio de Berríos (2005), el cual realiza un análisis sobre la vida cotidiana del cuerpo académico en el sistema universitario. En su análisis incorpora la metáfora del “principio de mateo”, según el cual la mujeres tendrían que alcanzar estándares de desempeño más alto que los hombres para igualarlos en prestigio. Por lo tanto nos encontraríamos en un contexto en el que las condiciones para realizar satisfactoriamente el trabajo académico no serían las mismas para las mujeres que tienen

que cumplir con patrones más altos de educación, capacitación y desempeño para competir con sus colegas varones, y así neutralizar el coeficiente simbólico negativo con el que son asociadas (Berríos, 2005:355).

De acuerdo Becher (2001) lo que se busca en el ámbito académico no es ni poder ni dinero como en el resto de la sociedad, sino lo que él llama la “buena reputación” (Becher, 2001). Esta es la misma idea que sostiene Bourdieu en “el oficio del científico” (2003), donde señala que en el campo académico lo que hay en el fondo es un intento por obtener renombre o reconocimiento profesional. La naturaleza de la actividad académica implica estar constantemente en busca de más prestigio, a la vez que revalidar el ya obtenido (Bourdieu, 2003).

En el mundo de la universidad, por lo tanto, las representaciones sobre lo femenino y lo masculino son la base sobre la cual se comienza a buscar ese prestigio. De manera tal que en un contexto que pide más a las mujeres para alcanzar puestos directivos, es necesario acumular mucho prestigio, lo que implica desarrollar arduamente tanto habilidades de gestión como de “peso académico”. Este último componente implica investigación, extensión, reconocimiento de los pares y desempeño en docencia (Lamadrid, 2012).

Sobre el desempeño en docencia ya habíamos establecido que uno de los indicadores más validados resulta ser la evaluación docente, por lo que la evaluación por parte de los estudiantes resulta fundamental como parte de acumulación y revalidación de prestigio en un ámbito académico. Pasaremos ahora por lo tanto a describir lo que se ha escrito sobre la relación entre el desempeño docente y el sesgo de género en el ámbito de la evaluación docente.

b.2. Sesgos de género en la evaluación docente

Relacionado al punto anterior, las investigaciones que tienen como objeto de estudio el género del docente se han ido incrementando a medida que ha ido aumentando la proporción de docentes mujeres en la educación superior. Hay que partir señalando que el grueso de la investigación reciente concluye efectivamente que ambos sexos son evaluados de forma distinta en relación a distintos ítems (Garduño, 2000). Sin embargo ha sido difícil

llegar a estas conclusiones sobre todo por las limitaciones que supone estudiar esta materia por medio de estudios cuantitativos.

En una investigación en torno al tema del sesgo de género realizada por Sprague y Massoni (2005), luego de una evaluación crítica sobre la literatura que ha tratado el tema, se concluyó que el sesgo impuesto por esta variable no puede ser fácilmente detectable por las escalas cuantitativas, agregando que la incapacidad que ha habido para pesquisar diferencias en cuanto al género se debe específicamente a un problema metodológico. En una escala numérica clásica, los investigadores por lo general no observan mayores diferencias de género, sin embargo afirman que estos análisis parten de dos premisas bastante cuestionables. Primero, se asume una métrica universal (que responder en el número 5, por ejemplo, tiene siempre el mismo significado, independiente del profesor); segundo, que el ranking que se les asigna a los profesores corresponde a conductas equivalentes. En efecto, los autores afirman a partir de las investigaciones precedentes se ha podido observar que los estudiantes utilizan unas líneas de base para evaluar a los hombres que se diferencian de las utilizadas a la hora de evaluar a las mujeres (Sprague y Massoni, 2005).

Para explorar esta posibilidad, los autores realizaron un análisis cualitativo de las palabras y adjetivos que 288 estudiantes universitarios, en dos campus, utilizan para describir al mejor y al peor profesor que han tenido. A su vez tuvieron que reportar su género, raza/etnia, y el año en que lo tuvieron. Aunque se encontró una considerable superposición en las formas en que los estudiantes hablaban de sus maestros y maestras, también se encontró indicios de que los estudiantes tienen a los profesores asociados a ciertas expectativas de género. Los adjetivos más comunes para describir a un buen profesor apuntaban más bien a lo que el profesor “es” (comprensivo, chistoso), mientras que en el caso de las profesoras apuntaba a los que ellas “hacen” (útil, atento). El siguiente paso fue agrupar los conglomerados de palabras en factores. La estructura de estos factores fue construida para cada género de forma separada. Los autores encontraron que las expectativas las cargan todos los profesores, pero la carga de las mujeres está más asociada a un trabajo intensivo. También se vieron señales de hostilidad hacia las mujeres mucho mayor que hacia los hombres que no cumplen las expectativas de los estudiantes (Sprague y Massoni, 2005).

Los investigadores Sprague y Massoni, (2005) concluyen que subyace a los métodos de evaluación aparentemente equivalentes un sesgo de género encubierto. Por tanto, se sugiere que las valoraciones de los estudiantes deben interpretarse en relación a comentarios cualitativos de los mismos, controlando el género de los estudiantes, e incluso entonces interpretarse con cautela.

Pese a las dificultades, mencionadas por los autores, para captar estas diferencias por medio de métodos cuantitativos, hay estudios que han arrojado resultados interesantes en esta línea. En un meta-análisis de estudios cuantitativos realizado por Feldman (1993), el autor encontró diferencias de género en la evaluación de los profesores por parte de los estudiantes en 28 de los 39 estudios analizados. En términos de dimensiones, las profesoras mujeres fueron mejor evaluadas en sensibilidad, preocupación, disponibilidad y ayuda.

Siguiendo esta línea, la socióloga Susan Basow (2000) confirmó los hallazgos de Feldman respecto a la correlación entre los puntajes y el género del docente. La autora se refirió también a la dificultad asignada por ciertos autores para encontrar diferencias de género por medio de métodos cuantitativos, y tomando como ejemplo el meta análisis de Feldman (1993), la autora dirá que cuando los descriptores se agrupan en factores o puntajes agregados, no aparecen efectos significativos de género, lo que sugiere que las variables de género suelen operar en el micro nivel, como descriptores específicos. La tendencia a analizar los resultados utilizando factores y puntajes agregados parece ser la razón por la cual los estudios cuantitativos globales con frecuencia no encuentran diferencias entre las puntuaciones de instructores hombres y mujeres.

En uno de sus estudios más importantes sobre esta materia, Basow (1995) realizó un estudio multivariado sobre la interacción entre el género del profesor y el género del alumno en función de los resultados obtenidos por las evaluaciones de los estudiantes en un período de 4 años en una universidad privada de artes liberales. Los resultados mostraron efectos significativos para cada uno de los 4 semestres examinados en dimensiones específicas que fueron desglosadas. Por un lado, las calificaciones de los profesores varones parecen no ser afectadas por el sexo del estudiante. Por el contrario, las profesoras mujeres tendían a recibir sus calificaciones más altas de estudiantes mujeres y sus calificaciones más bajas de los estudiantes varones. Las calificaciones promedio recibidas por los

profesores hombres y mujeres también varían en función de la facultad y de la disciplina académica (1995). En este sentido, se encontró que los profesores hombres en ciencias sociales eran mejor evaluados que las profesoras en dimensiones afectivas como: entusiasmo, estimulación y retroalimentación. También se observó diferencias en los puntajes obtenidos en Ciencias donde la evaluación general de los hombres fue más alta que la de las mujeres.

En un estudio más reciente de la misma autora y colaboradores (Basow, Phelan y Capotosto, 2006) realizado en una Universidad pública de California, corrobora esta hipótesis donde como era de esperar, los factores de género jugaron un rol fundamental en la evaluación de los mejores y peores profesores por parte de los estudiantes. Como ya lo habían demostrado estudios previos de esta autora (Basow, 2000; Basow 1995), los estudiantes masculinos eran particularmente propensos a ocupar estereotipos de género al momento de evaluar a profesoras mujeres. Los puntajes que los estudiantes hombres asignaron a sus profesoras fueron significativamente inferiores de lo esperado en relación al porcentaje de profesores mujeres que tuvieron en clases. Esto sugiere que los estudiantes hombres tendrían más actitudes tradicionales hacia las mujeres, por lo que califican a las profesoras más bajo que las estudiantes mujeres. Las estudiantes de sexo femenino, por su parte, evaluaron a sus profesores como buenos y malos en proporción directa a su frecuencia en el aula. Por lo tanto, las estudiantes mujeres no parecen estar tan determinadas por el género como un factor principal en su elección. Así mismo, las descripciones de los mejores y peores profesores variaban por el profesor y el género del estudiante. En general, la pareja: profesor femenino - estudiante masculino, parecía estar más afectado por las dinámicas de género, siendo los estudiantes varones más propensos a enfatizar habilidades interpersonales, sobre todo el ser accesible. Los estudiantes varones también eran más propensos a describir como peor profesor a las mujeres en términos de las interacciones en clases, especialmente el ser “cerrada de mente”.

Sobre lo anterior hay que matizar con dos aspectos fundamentales. Primero, que en áreas como las humanidades, ciencias sociales y comunicaciones, la mayoría de los estudios han demostrado que estudiantes hombres y estudiantes mujeres tienden a evaluar a sus profesoras mujeres mejor que a sus profesores hombres en general (Tatro, 1995; Basow,

1998; Smith et al 2007). Respecto a diferencias en cuanto al género del estudiante, en el meta análisis de Aleamoni y Hexner (1980), los autores citan varios estudios de los años sesenta y setenta que llegaron a la conclusión de que las estudiantes mujeres tienden a evaluar mejor que los estudiantes hombres. En segundo lugar, hay que señalar que si bien estas diferencias existen, los tamaños de los efectos suelen ser muy pequeños. El hecho de que no se esperen diferencias fuertes tiene relación también con lo que ya señalábamos en cuanto a que el sesgo de género no es fácilmente percible dado el efecto ocultador que tienen los instrumentos cuantitativos (Sprague y Massoni, 2005). Sin embargo, el que el tamaño del efecto sea pequeño no resta que este pueda ser significativo, tanto por el tamaño de la muestra como por la consistencia de los datos (Smith, et al, 2007: 64).

La socióloga Susan Basow (1998) sugiere que más que esperar grandes tamaños de efectos estadísticos, los cuales por lo general son pequeños y suelen favorecer a las mujeres en términos generales, lo que interesa es analizar los efectos de interacción entre las variables de género y otras variables contextuales, ya que allí, en el ámbito específico, es donde se pueden analizar los patrones de género de manera más sustantiva, y es ahí donde interesa que los hallazgos sean significativos y consistentes entre sí, pues ello indica relaciones, que aunque sean de fuerza menor, son teóricamente relevantes desde la perspectiva de género.

Capítulo 3: Marco Metodológico.

I. Hipótesis de investigación

Hipótesis General

Los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, están asociados significativamente con sesgos relativos al género del profesor y del alumno y este sesgo varía significativamente según el año de estudios cursado.

Hipótesis Específicas

1. Los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile varían significativamente en relación al ciclo de estudios cursado por el alumno.
2. Los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, varían significativamente en relación al género del alumno.
3. Los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, varían significativamente en relación al género del profesor.
4. Los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, varían significativamente en relación a la interacción entre el género del profesor, el género del alumno y el nivel o año cursado por el alumno.

II. Metodología de investigación

La presente investigación posee un carácter no experimental, dado que no se controlaron variables ni las condiciones de experimentación y está constituida desde un aspecto exclusivamente cuantitativo correlacional. De esta manera, el diseño de investigación del presente estudio se plantea como una metodología cuantitativa correlacional no experimental.

Los datos cuantitativos se obtuvieron a partir del cuestionario de evaluación docente que respondieron estudiantes de la facultad de ciencias sociales de la Universidad de Chile el año 2007.

A. Técnica de recolección de datos: Cuestionario de desempeño docente

Para producir los datos cuantitativos se utilizó un cuestionario o encuesta de evaluación de desempeño docente el cual se intentó aplicar al total de estudiantes de la facultad de ciencias sociales de la Universidad de Chile el año 2007. En este sentido, se utilizó un enfoque censal y no muestral.

La recolección de datos se realizó por medio de una escala Likert. Esta es una escala psicométrica comúnmente utilizada en cuestionarios, y es la escala de uso más amplio en encuestas para la investigación, principalmente en ciencias sociales.

Esta escala se basa en un método de escalamiento centrado en los sujetos, donde las variaciones sistemáticas encontradas en las reacciones de los sujetos ante los estímulos se atribuyen a diferencias individuales, lo que permite situar a los individuos a lo largo de un continuo psicológico. Lo que se busca, en el fondo, es que se manifiesten las diferencias individuales (Barbero, 2003).

Los fundamentos o principios de la escala Likert siguiendo a Barbero (2003) serán:

1. Es posible medir diversas dimensiones de una actitud a partir de un conjunto de enunciados que operen como reactivos para los sujetos.
2. La variación de las respuestas será producto de las diferencias individuales de los sujetos.

3. La valoración de los sujetos en la variable no supone una distribución uniforme sobre el continuo de la actitud, sino una posición favorable o desfavorable sobre ese objeto estudiado.

B. Universo y muestra del Estudio

El universo serán todos los estudiantes de pregrado de la Facultad de ciencias sociales Universidad de Chile el año 2007.

La muestra es la información real disponible para este estudio que fue de 9132 cuestionarios respondidos anónimamente por cada estudiante, lo que corresponde a la evaluación que realizó cada alumno de cada uno de los profesores que tuvo en cada uno de sus cursos en los dos semestres académicos del año 2007.

Si bien se intentó obtener las respuestas del total de estudiantes de la Facultad respecto de la totalidad de sus docentes, por diversos motivos (enfermedad, abandono del semestre, ausencias no justificadas, etc), algunos estudiantes no contestaron el instrumento. No obstante, se obtuvo las respuestas del 83% del universo, lo que implica que la información disponible muy probablemente es un buen indicador del total de la opinión de los estudiantes de la Facultad de Ciencias Sociales.

C. Operacionalización de las variables

Como variables dependientes se considerarán los distintos ítems, dimensiones y puntuación total del cuestionario, mientras que como variables independientes utilizamos los llamados factores extra-clase definidos previamente en el marco teórico. Hemos seleccionado dos variables correspondientes a características propias del alumno (semestre cursado y género del estudiante), y otra correspondiente a características del profesor (género del profesor). Se buscará ver la interacción de estas variables independientes y su incidencia en la variable dependiente, que será el puntaje otorgado por los alumnos al profesor en cada pregunta y en las dimensiones de la encuesta de evaluación docente.

- **Variables Dependientes**

Puntaje obtenido por los profesores en la evaluación docente → se refiere al puntaje promedio obtenido por los profesores en cada uno de los ítems o variables. Estos estarán diferenciados por género del profesor y del alumno. Los puntajes están expresados en una escala de 5 puntos.

Las variables incluidas en el estudio serán agrupadas en dos factores. El procedimiento del análisis factorial confirmatorio lo detallaremos en los resultados del análisis, pero por ahora interesa exponer simplemente las variables incluidas en el estudio agrupadas por factores.

Tabla N°1: Variables dependientes

Factor	Ítem / indicador
Factor 1 :Desempeño en el aula	4.- El profesor domina los contenidos de su clase.
	5.- El profesor se comunica con claridad (al responder preguntas, explicar contenidos, etc.).
	6.- Los contenidos expuestos por el profesor están actualizados y tienen relación con la realidad actual.
	7.- El profesor presentó variados enfoques teóricos referentes a la materia.
	8.- El profesor utiliza estrategias de enseñanza que estimulan la participación de los estudiantes.
	9.- El profesor demuestra compromiso con el desarrollo del aprendizaje de los estudiantes.
	10.- El profesor muestra disposición a enseñar y aclarar dudas, tanto dentro como fuera de la hora de clases.
	11.- El profesor acerca al estudiante al conocimiento de la disciplina a través de situaciones concretas o ejemplos.
	12.- El profesor logra un trato equitativo y respetuoso en su relación con los estudiantes.
	13.- El profesor es consistente entre lo enseñado y lo exigido en las evaluaciones.
Factor 2: Responsabilidad	1.- El profesor inició y terminó puntualmente su clase.
	2.- El profesor fue puntual en la entrega de evaluaciones.
	3.- El profesor asistió regularmente a clases.
Ítems compartidos	15.- El profesor entrega el programa oportunamente (al inicio del semestre).
	16.- El profesor entregó pautas claras de corrección de pruebas y trabajos.

El sentido de trabajar tanto con puntuaciones agrupadas en factores como con ítems aislados tiene relación con lo encontrado en estudios previos respecto a que si bien podría no detectarse un sesgo de género en el conjunto de las respuestas, es posible que este sesgo se evidencie en las afirmaciones puntuales y en sus interacciones específicas.

- **Variables Explicativas (o “Independientes”)**

Género del alumno → Se refiere al sexo del alumno que responde cada cuestionario. En este caso, de acuerdo a los objetivos del estudio, se remitirá únicamente a dos posibilidades: hombre o mujer. Este indicador nos permitirá relacionar esta variable con el puntaje obtenido, así como ver su interacción según año de estudio y género del profesor.

Género del profesor → Se refiere al sexo del profesor que dictó cada uno de los cursos. En este caso, de acuerdo a los objetivos del estudio, se remitirá únicamente a dos posibilidades: hombre o mujer. Este indicador nos permitirá relacionar esta variable con el puntaje obtenido, así como ver su interacción según año de estudio y género del alumno.

Ciclo o nivel de estudios → Se refiere al nivel de estudios que está cursando el estudiante en el momento de responder el cuestionario. Basado en la idea de la formación por competencias que separa entre dos ciclos en la formación universitaria: básico y avanzado, los estudiantes se agruparán en dos categorías.

- Cursando etapa básica (1° y 2° año)

- Cursando etapa de especialización o avanzada (3°, 4° y 5° año)

Este indicador nos permitirá ver si los estudiantes del primero o segundo año están evaluando mejor o peor a sus profesores que los estudiantes de los últimos años. De la misma manera se analizará la interacción con el resto de las variables independientes del estudio y cómo afectan en conjunto la a variable dependiente.

III. Plan de Análisis de la Información

En términos generales, la investigación se realizó utilizando técnicas de análisis estadístico bivariado y multivariado. Al manejar variables independientes y dependientes, la investigación fue de carácter asociativa y tomó como base la información producida por los cuestionarios aplicados a los estudiantes de pregrado de la Universidad de Chile.

El análisis de la información se realizó en distintas fases. Para sintetizar la exposición de los resultados, no toda la información será reportada en el cuerpo del estudio, pero si podrá ser consultada en los anexos.

A. Prueba de normalidad de las variables:

Para saber qué técnica de análisis se iba a utilizar, lo primero que se hizo fue saber si los datos se distribuyen de manera normal. Para ese propósito se utilizaron medidas de distribución que permitieron identificar la forma en que se separan o aglomeran los valores. Las medidas utilizadas fueron la Asimetría y la Curtosis.

Las medidas de asimetría buscan establecer el grado de simetría (o asimetría) de una distribución entendiendo por simetría la distancia con respecto al eje determinado por la media aritmética. Por tanto, diremos que una distribución es simétrica cuando los valores de la variable equidistantes de este valor central tengan la misma frecuencia, en caso contrario diremos que es asimétrica, siendo esta asimetría negativa si es más larga la rama de la izquierda o positiva si la rama de la derecha es más larga. Desde luego entre mayor sea el número (positivo o negativo), mayor será la distancia que separa la aglomeración de los valores con respecto a la media.

Respecto a la curtosis o apuntamiento, esta medida nos dirá el grado de concentración que presentan los valores en la región central de la distribución. Por medio del Coeficiente de Curtosis, podemos identificar si existe una gran concentración de valores (Leptocúrtica), una concentración normal (Mesocúrtica) ó una baja concentración (Platicúrtica).

Por tanto, revisamos la distribución de cada ítem por medio de estas tres medidas comprobando que se comportan dentro de los márgenes de normalidad, tras lo cual pudimos continuar el análisis (Revisar Anexo II).

B. Fase de diseño de las variables dependientes

Con el fin de disponer de puntuaciones para las diversas dimensiones que mide el cuestionario de evaluación docente, se realizó un análisis factorial exploratorio y confirmatorio que agrupó los ítems en dimensiones.

El análisis factorial en términos generales consiste en un conjunto de técnicas multivariantes de interdependencia que se utilizan con el propósito de estudiar los patrones de interrelación de un conjunto numeroso de variables manifiestas (observadas) mediante la generación de un conjunto más reducido de variables latentes (no observadas). Esta técnica nos permite respaldar o no la consistencia interna de la escala, es decir, si existe o no consistencia entre lo que la teoría plantea acerca de las dimensiones de la escala y lo que empíricamente se constata acerca de estas dimensiones.

El análisis factorial exploratorio se realizó por medio del programa Factor Analysis 8.02 (Lorenzo-Seva y Ferrando, 2011), que es un programa especial para realizar análisis factorial con variables ordinales, como las nuestras, utilizando correlaciones tetracónicas. El análisis factorial confirmatorio se realizó por medio del programa LISREL 8.8 (Jöreskog & Sörbom 2006) diseñado para trabajar ecuaciones estructurales y modelos de medida con variables latentes.

c. Análisis univariante

Con el fin de determinar el grado de asociación entre las variables independientes y dependientes se realizó un análisis T de diferencia de medias. La prueba T de igualdad de medias sirve para ver si existen diferencias entre los grupos definidos por las variables independientes dicotómicas (sexo del alumno, sexo del profesor y nivel de estudios) y cada una de las variables dependientes correspondientes a cada ítem de la evaluación, a los que sumamos los dos factores generales. Por lo tanto se realizaron 54 pruebas de diferencia de medias.

Se evaluó el comportamiento de las diferentes variables escogidas y la capacidad explicativa de cada una de ellas.

Se utilizó como criterio de significación $p < 0,05$, pero dado que se trabajará con una muestra de tamaño muy grande, posiblemente efectos de muy pequeña magnitud serán significativos, por lo que esta información se complementará con el cálculo de los tamaños del efecto por medio del estadístico D de Cohen y R de Pearson.

d. Tamaño del efecto en análisis univariante.

Interesa revisar el tamaño del efecto básicamente por el hecho de que una diferencia estadísticamente significativa no es necesariamente una diferencia grande y tampoco es necesariamente una diferencia importante.

En ese sentido, un valor estadísticamente significativo de t nos permite afirmar que la diferencia entre las medias de las poblaciones representadas por esas dos muestras no es cero (con más propiedad, que es muy improbable que sea cero) pero no hay relación entre el valor de t y la magnitud de la diferencia, porque el valor de t no depende solamente de la diferencia entre las dos medias, sino de las varianzas de las muestras y sobre todo del tamaño de las muestras.

Por eso es que el concluir una investigación rechazando sin más la hipótesis nula ha sido muy criticado por numerosos autores (Morales, 2012). Una forma de resolver ese problema es calcular la magnitud o tamaño del efecto.

El tamaño del efecto, como la misma palabra tamaño expresa, nos va permitir hablar de magnitudes, de diferencias grandes o pequeñas y consiguientemente de la relevancia de la diferencia encontrada. Su fórmula es bastante básica y consiste en la diferencia de las medias observadas dividida por la desviación estándar.

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sigma}$$

Por otro lado también se usa como tamaño del efecto transformar los resultados de las pruebas estadísticas a r de Pearson. Su fórmula está basada en los resultados obtenidos de la prueba anterior.

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

Respecto a cuándo se puede considerar grande o pequeño un determinado valor, suelen aceptarse las orientaciones ofrecidas por Cohen (1988):

Acá señala como criterios:

- a. Las diferencias pequeñas (en torno a $d = .20$): Cohen dirá que pueden parecer muy pequeñas como referencia útil, pero estas pequeñas diferencias no deben descartarse ya que pueden igualmente ser de interés en áreas nuevas de investigación, en estudios exploratorios, o cuando no se controlan bien otras variables, etc. En el fondo son diferencias lo suficientemente relevantes como para ser consideradas en un estudio.
- b. Diferencias moderadas (en torno a $d = .50$) se entienden aquí aquellas diferencias entre grupos que pueden detectarse por simple observación.
- c. Diferencias grandes a partir de $d = .80$

Si bien estas orientaciones son un tanto arbitrarias, son muy aceptadas como razonables y se citan frecuentemente en la bibliografía experimental, aun así no las tomaremos como norma absoluta, pero sí como orientación.

Sin embargo, se ha señalado también que cuando se trata de resultados de rendimiento escolar o de investigación educacional suele considerarse valores bastante menores (en torno a $.30$), como de significación práctica o importante (Morales, 2012). La principal razón es simplemente que en educación suelen encontrarse tamaños del efecto menores que en otras disciplinas. Esto es así producto de que los criterios de Cohen están pensados para estudios experimentales, en que la diferencia entre los dos grupos era producida por una intervención del investigador y por tanto podía ser muy fuerte, y no estaba pensando en estudios transversales por encuestas.

Sobre esto escribieron Valentine y Cooper (1993), señalando que a diferencia de los estudios de gran escala en los que pensaba Cohen, en estudios orientados al ámbito educacional valores en torno a $.30$ se juzgan ya relevantes (Valentine y Cooper, 2003).

De esta forma mantendremos la tipología de Cohen, es decir, tomando en consideración los tamaños del efecto cercanos a 0.20 pero considerando ya de efecto fuerte los cercanos al 0.30.

Finalmente, emplearemos la sugerencia de Morales (2011), quién señala que otra manera de valorar los tamaños del efecto es compararlos con otros obtenidos en términos relativos dentro del mismo estudio, por lo que también usaremos los tamaños del efecto para comparar variables y así ver dónde hay mayores o menores diferencias.

d. Análisis de varianza factorial

Para evaluar los efectos individuales y conjuntos realizaremos un análisis de varianza factorial. Este análisis sirve para evaluar la capacidad explicativa de un modelo que puede incluir los efectos principales (independientes) de más de una variable independiente categórica (factores), y las interacciones entre dichos efectos principales, sobre una variable dependiente cuantitativa.

En este caso, el análisis de ANOVA factorial nos permitirá estudiar, si las puntuaciones obtenidas en los distintos ítems y en los dos factores (variables dependientes) dependen de: i) sexo de los estudiantes que evalúan (primera variable independiente); ii) sexo del profesor (segunda variable independiente); iii) el ciclo de estudios (tercera variable independiente). También podremos estudiar el poder explicativo de las interacciones ya que podremos observar si las diferencias entre las evaluaciones de estudiantes hombres y mujeres se mantienen constantes según el sexo del profesor o según el ciclo de estudios, es decir, permite determinar si la interacción entre los factores sexo del alumno, sexo del profesor y nivel de estudios afecta a la variable dependiente de cada ítem.

En cuanto a la especificación del modelo realizaremos un análisis factorial completo, ya que de esta forma podremos ver las interacciones de todas las variables. En este tipo de modelos los efectos de interés son: los tres efectos principales, los tres efectos de las interacciones dobles (uno por cada interacción entre cada dos factores) y el efecto de la interacción triple (entre los tres factores).

Lo anterior es análogo a realizar una regresión con efectos de interacción, donde la prueba evalúa cuanto explica cada variable por separado, y luego, de la varianza que queda sin explicar, evalúa cuanto es lo que se explica producto de las interacciones entre dos o más variables. Es un modelo Multi-etápico, lo que anula básicamente el riesgo de multicolinealidad. Es una regresión por etapas de complejidad, ya si se puede explicar algo con los efectos principales, no será necesaria ni relevante la interacción, por lo que el modelo está basado en el principio de parsimonia.

Es importante señalar que en un análisis de varianza factorial existe una hipótesis nula por cada variable independiente y por cada posible combinación de ellas:

- La hipótesis nula relacionada a cada factor fijo dirá que las medias de las poblaciones definidas por los niveles del factor son iguales.
- La hipótesis nula relacionada al efecto de una interacción afirma que tal efecto es cero.

Para contrastar estas hipótesis, el ANOVA factorial se sirve de estadísticos F, donde el nivel crítico asociado a cada estadístico es lo que nos permite decidir si podemos mantener o debemos rechazar la hipótesis nula. Para esto usaremos un nivel de significación $p < 0.05$. Respecto a las condiciones de aplicación, podemos decir que el modelo supone que los grupos o poblaciones obtenidas, definidas por la cantidad de categorías de las variables independientes, son normales (supuesto de normalidad) y que sus varianzas son iguales (supuesto de homocedasticidad).

Respecto al primer supuesto, este ya ha sido revisado en secciones previas mediante test de normalidad de las variables.

Respecto a la homocedasticidad, usaremos principalmente como criterio de adecuación del modelo logrado los gráficos de residuos tipificados. Este nos señalará si los residuos tipificados tienen o no relación con los valores predichos, donde lo que se espera es que los residuos esperados y tipificados formen nubes sin forma. Esto se realizará para cada variable y así podremos corroborar la igualdad de varianzas necesaria para continuar el análisis.

En cuanto a los descriptivos que obtendremos de estas interacciones, interesará sobre todo el estadístico del tamaño del efecto (η^2 cuadrado). A través éste podremos medir la intensidad de la asociación entre ciertas variables. Es decir, del grado en que cada factor o combinación de factores está afectando a la variable dependiente.

El estadístico η^2 cuadrado parcial, se obtiene para un efecto concreto por medio de la división de la Suma de Cuadrados entre grupos y la Suma de Cuadrados total. Este estadístico se interpreta como proporción de varianza explicada: es una estimación de la proporción de variación de la variable dependiente que está explicada por cada efecto.

En los casos en que existan efectos significativos de la interacción se obtendrán las medias marginales estimadas para interpretar mejor esos efectos. En segundo lugar pediremos los gráficos de perfil. Hay que entender que los efectos de interacción son efectos que cambian siempre las pendientes de las curvas, por lo tanto cuando las líneas de ambas categorías se mantienen paralelas, no hay efectos de interacción. La mejor muestra de que estamos en presencia de efectos de interacción es cuando las líneas están cruzadas, por lo que se entiende que los efectos de interacción son fuertes.

Finalmente, como este análisis está pensado para variables dependientes continuas, y nuestras variables dependientes son ordinales, corroboramos estos análisis con tablas de chi-cuadrado múltiples y el estadístico Gamma, que son estadísticos específicamente para variables ordinales. Estos análisis no son reportados, sino que se hicieron para confirmar lo que ya habíamos obtenido con el análisis de varianza factorial.

Capítulo 4: Resultados y análisis

I. Análisis factorial

A. Análisis exploratorio: resultados del programa Factor.

Comenzamos con un cuadro de detalle del análisis y el procedimiento (Tabla N°2). Se utilizaron correlaciones tetracóricas o policóricas. El método de extracción será el Unweighted Least Squares y el de rotación es el Promin. Finalmente el procedimiento para determinar el número de factores fue el de implementación óptima por medio del análisis paralelo. Este análisis mostró una solución ajustada con dos factores. Una vez verificado en base al coeficiente de asimetría (Skewness) y al coeficiente de Mardia, que la distribución es Simétrica, se procedió a rotar la matriz. Podemos ver que si efectivamente el análisis se rota por dos factores hay 5 variables que son mejor explicadas por el segundo factor.

Tabla N° 2: Matriz rotada

ROTATED LOADING MATRIX (loadings lower than absolute 0.300 omitted)		
Variable	F 1	F 2
V 1		0.917
V 2		0.914
V 3		0.892
V 4	0.686	
V 5	0.865	
V 6	0.814	
V 7	0.754	
V 8	0.861	
V 9	0.845	
V 10	0.839	
V 11	0.884	
V 12	0.780	
V 13	0.699	
V 14	0.615	
V 15		0.553
V 16	0.341	0.459

EXPLAINED VARIANCE AND RELIABILITY OF ROTATED FACTORS Mislevy & Bock (1990)		
Factor	Variance	Reliability estimate
1	7.339	0.961
2	3.364	0.900

b. Análisis confirmatorio

El análisis exploratorio realizado con Factor nos indicó la posible existencia de dos factores, uno mayor que incluía un mínimo de 12 ítems y uno menor con entre 3 a 5 ítems (lo que implica que los factores podían compartir uno o dos ítems). En base a ello se intentó confirmar esta estructura por medio de un análisis factorial confirmatorio realizado con el programa LISREL. Pusimos a prueba 5 modelos.

El primer modelo constituye un modelo con un solo factor para los 17 ítems. El segundo implicó dos factores, estando el primero formado por 14 ítems y el segundo sólo por 3. El tercer modelo consideraba un primer factor con 12 ítems y un segundo con 5. El cuarto modelo implicó suponer que el primer factor constaba de 13 ítems y el segundo 5, con lo que ambos comparten un factor. Finalmente el quinto modelo implicó suponer un primer factor que explicaba 14 ítems y un segundo con 5 ítems, compartiendo ambos 2 ítems.

Tabla N° 3: Ajustes de los Modelos (con correlaciones tetracóricas)

Modelo	Chi² (gl)	DIF Chi²	RMSEA	SRMR	AGFI	NNFI	CFI
1 Factor	20084 (104)	-----	0.16	0.066	0.99	1.00	1.00
2 factores (14I/3I)	11936 (103)	8148 (1)	0.13	0.048	0.99	1.00	1.00
2 factores (12I/5I)	11815 (103)	121 (0)	0.13	0.044	0.99	1.00	1.00
2 factores (13I/5I, 1 comp.)	10573 (102)	1242 (1)	0.12	0.040	0.99	1.00	1.00
2 factores (14I/5I, 2 comp.)	10033 (101)	540 (1)	0.12	0.036	1.00	1.00	1.00

Vemos que el Chi² del primer modelo es cercano a 20 mil, con 104 grados de libertad. Por convención la significación deberá ser igual a la cantidad de grados de libertad, por lo que en el primer modelo claramente se rechaza la hipótesis nula. Al ver los otros modelos ocurre exactamente lo mismo. En rigor, para la cantidad de datos ninguno de estos modelos se ajusta realmente. Sin embargo, lo sustantivo es saber si el grado de desajuste es relevante o irrelevante. Las siguientes columnas nos podrán dar mayores luces al respecto.

La diferencia de Chi cuadrado es la diferencia de un modelo con el anterior. Es la resta por ejemplo del primer modelo por el segundo, además restamos los grados de libertad. La

diferencia de los modelos muestra que ésta si es significativa, ya que con solo 1 grado de libertad la significación es de 8 mil.

Los modelos posteriores sin embargo caen bastante en cuanto a su nivel de significación. De esta manera, con esta jerarquía anidada de modelos cada vez más complejos podemos escoger de mejor manera con qué modelo nos quedaremos.

Los siguientes indicadores prueban la validez de nuestros modelos por medio de pruebas de bondad de ajuste relativos.

El RMSEA, supone por convención que se debe obtener una puntuación de 0.08 o menos para que exista ajuste. Incluso para nuestro modelo más complejo, este indicador señala que nuestro modelo no logra ajustar del todo. Sin embargo, este es un tipo de indicador que suele tener problemas cuando los lambda (pesos factoriales) son muy altos, cosa que ocurre en nuestro estudio.

El SRMR, debe ser, por convención, menor que 0,05 para que se hable de buen ajuste. En el último modelo vemos que nos dio 0.036. Por lo que entendemos que el modelo deja residuales estandarizados pequeños. Los últimos 3 índices, AGFI, NNFI y CFI, nos pide por convención un valor de 0.95 y en nuestro modelo tenemos un 1.00, por lo que podemos confirmar el buen ajuste de nuestros datos.

La conclusión es que un modelo de 2 factores es mejor que el de 1. Y dentro del de 2 factores el que se muestra como más completo será el de 12 ítems en un factor y 5 ítems en el otro, con 2 de ellos Bi factoriales. Para resolver esto último construiremos factores con sumatoria ponderada. De esta manera dividiremos su varianza en ambos factores en función de los valores de los pesos factoriales.

Respecto a la distribución y coherencia de los ítems agrupados en cada factor, los resultados nos muestran que el primer factor agrupa un abanico bastante amplio que va desde ítems más relacionados al conocimiento y dominio de los contenidos, hasta ítems más relacionados con la empatía y de trato hacia los estudiantes. Sin embargo son coherentes entre sí tomando en cuenta que todos refieren al desempeño del profesor dentro del aula. Es decir, tienen relación con la capacidad pedagógica y docente del profesor y no

con aspectos más formales o externos a lo que ocurre en el aula. Por eso al primer factor se le llamó “Desempeño en el aula”.

El segundo Factor dirá relación justamente con aquellos ítems que no tienen tanta relación con el desempeño pedagógico mismo, sino con aspectos más formales y objetivos asociados al cumplimiento del reglamento. Por esa coherencia interna del factor se le llamó “Responsabilidad”

Finalmente los ítems compartidos (“el profesor entrega el programa” y “el profesor entregó pautas claras de corrección”) y que ponderan en ambos factores tendrán coherencia teórica también con ambos factores. Esto es así ya que si bien son ítems relacionados principalmente con aspectos de responsabilidad, no son tan objetivos como los ítems que agrupa el factor dos, ya que requieren una elaboración por parte del docente, ya sea presentando y diseñando el programa o bien haciendo y exponiendo una pauta de corrección clara y precisa.

Los ítems quedan distribuidos en 2 factores de la siguiente manera:

Tabla 4: Distribución en los factores

Factor	Ítem / indicador
Factor 1: Desempeño en el aula	4.- El profesor domina los contenidos de su clase.
	5.- El profesor se comunica con claridad (al responder preguntas, explicar contenidos, etc.).
	6.- Los contenidos expuestos por el profesor están actualizados y tienen relación con la realidad actual.
	7.- El profesor presentó variados enfoques teóricos referentes a la materia.
	8.- El profesor utiliza estrategias de enseñanza que estimulan la participación de los estudiantes.
	9.- El profesor demuestra compromiso con el desarrollo del aprendizaje de los estudiantes.
	10.- El profesor muestra disposición a enseñar y aclarar dudas, tanto dentro como fuera de la hora de clases.
	11.- El profesor acerca al estudiante al conocimiento de la disciplina a través de situaciones concretas o ejemplos.
	12.- El profesor logra un trato equitativo y respetuoso
	13.- El profesor es consistente entre lo enseñado y lo exigido en las evaluaciones.
	14.- El profesor prepara materiales útiles y pertinentes para el aprendizaje.

Factor 2: Responsabilidad	1.- El profesor inició y terminó puntualmente su clase.
	2.- El profesor fue puntual en la entrega de evaluaciones.
	3.- El profesor asistió regularmente a clases.
Ítems compartidos	15.- El profesor entrega el programa oportunamente (al inicio del semestre).
	16.- El profesor entregó pautas claras de corrección de pruebas y trabajos.

II. Análisis Univariado

A continuación se presenta un resumen de los resultados de la prueba T de igualdad de medias, realizada para ver si existen diferencias entre los grupos definidos por las variables independientes dicotómicas (sexo del alumno, sexo del profesor y nivel de estudios) y cada una de las variables dependientes correspondientes a cada ítem de la evaluación más los dos factores construidos mediante el procedimiento factorial expuesto anteriormente. Por lo tanto se realizaron 54 pruebas de diferencia de medias. Estos resultados se sintetizaron luego de revisar la prueba de igualdad de varianzas e interpretar de acuerdo a ello el estadístico t con su nivel de significación bilateral.

La dirección de la asociación se indica en la columna “Puntaje mayor”, que indica que la media obtenida por aquella categoría de la variable independiente resultó ser la más alta en relación al puntaje de evaluación docente.

a. Género del alumno

La tabla 5 nos muestra que para todos los ítems hay diferencias de medias estadísticamente significativas, excepto para los ítems “Puntualidad entrega evaluaciones” y para el ítem “Entrega programa oportuna”. En ambos casos no hay diferencia de medias significativas. También se encontraron diferencias estadísticamente significativas para los dos factores “responsabilidad” y “desempeño”.

Los límites del intervalo de confianza para la diferencia fueron utilizados como complemento para corroborar si podemos o no rechazar la hipótesis nula de igualdad de medias. Este complemento se correspondió siempre con la prueba de significación por lo que los resultados son los mismos.

Respecto a la direccionalidad, en todos los casos se observó que las estudiantes mujeres tienen medias más altas que los estudiantes hombres, por lo que se puede concluir que las estudiantes mujeres evalúan mejor que los estudiantes hombres a sus profesores.

Tabla 5: Género del alumno¹

	Puntaje Mayor	D Cohen	R Pearson	t	gl	Sig.
Puntualidad inicio - término clases	Alumno (M)	-0,08	-0,04	-3.557	7783	0.00
Puntualidad entrega evaluaciones	Alumno (M)	-0,04	-0,02	-1.647	7627	0.10
Asistencia regular a clases	Alumno (M)	-0,10	-0,05	-4.487	7739	0.00
Dominio contenidos	Alumno (M)	-0,11	-0,05	-4.330	4.732.081	0.00
Comunicación clara	Alumno (M)	-0,15	-0,07	-6.676	7778	0.00
Actualización y relación con realidad actual	Alumno (M)	-0,17	-0,08	-7.298	7616	0.00
Presentar variados enfoques	Alumno (M)	-0,14	-0,07	-5.504	4.425.213	0.00
Estimula participación	Alumno (M)	-0,15	-0,07	-6.217	4.976.093	0.00
Compromiso con aprendizaje	Alumno (M)	-0,18	-0,09	-7.180	7737	0.00
Disposición a enseñar	Alumno (M)	-0,15	-0,07	-6.814	7683	0.00
Acerca conocimiento mediante ejemplos	Alumno (M)	-0,16	-0,08	-7.046	7684	0.00
Trato equitativo	Alumno (M)	-0,15	-0,07	-6.733	7767	0.00
Consistencia entre enseñado y exigido	Alumno (M)	-0,06	-0,03	-2.797	7647	0.05
Materiales útiles	Alumno (M)	-0,10	-0,05	-4.011	4.857.461	0.00
Entrega programa oportuna	Alumno (M)	-0,01	-0,01	-.441	7543	0.65
Entrega pautas corrección	Alumno (M)	-0,12	-0,06	-5.060	4.888.676	0.00
Factor 1 Desempeño	Alumno (M)	-0,17	-0,08	-6.639	6639	0.00
Factor 2 Responsabilidad	Alumno (M)	-0,09	-0,04	-3.763	4.821.799	0.00

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics y de Ncalculators: (utilizado el 9 de junio del 2012 en: <http://es.ncalculators.com/>)

Respecto a los tamaños del efecto, basados en la tipología de Cohen, es decir, tomando en consideración los tamaños del efecto cercanos a 0.20, o 0.10 en R de Pearson, vemos que la variables “compromiso con el aprendizaje” es la que resulta tener un mayor tamaño de efecto, seguida en orden de importancia por “actualización y relación con la realidad

¹ Para alumno hombre: Alumno (H);
Para alumno mujer: Alumno (M)

actual”, el factor “Desempeño”, “acerca el conocimiento mediante ejemplos” y “comunicación clara” (ver tabla N°6).

El tamaño del efecto de los primeros 3 ítems son bastante considerables y podrían considerarse de importancia práctica. Pero la conclusión más intuitiva a partir de estos datos es la consistencia de ellos, es decir, que la direccionalidad, independiente de más o menos tamaño del efecto, es siempre la misma.

Tabla 6: Cinco variables con mayor tamaño del efecto

	D Cohen	R Pearson
Compromiso con aprendizaje	-0.18	-0.09
Actualización y relación con realidad actual	-0.17	-0.08
Factor 1 Desempeño	-0.17	-0.08
Acerca conocimiento mediante ejemplos	-0.16	-0.08
Comunicación clara	-0.15	-0.07

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics y de Ncalculators: (utilizado el 9 de junio del 2012 en: <http://es.ncalculators.com/>)

b. Género del profesor

A diferencia de la variable género del alumno, en el género del profesor no hay una tendencia tan clara, ya que existen ciertos ítems en los cuales los profesores de sexo masculino son mejor evaluados y otros, la mayoría, en los que se evalúa mejor a los de sexo femenino.

Los ítems que no mostraron diferencia de medias significativas fueron “Comunicación clara”, “Estimula participación”, “Disposición a enseñar”, “Acerca conocimiento mediante ejemplos”, “Consistencia entre enseñado y exigido”, “Entrega programa oportuna”, y “Factor 1 desempeño”.

Respecto a la direccionalidad, en ocho ítems las profesoras mujeres resultaron ser mejor evaluada que los profesores hombres, estos son “Puntualidad inicio - término clases”, “puntualidad entrega evaluaciones”, “asistencia regular a clases”, “compromiso con aprendizaje”, “disposición a enseñar”, “trato equitativo”, “materiales útiles”, “entrega pautas corrección”, “Factor 2 Responsabilidad”.

Los profesores hombres por su parte fueron mejor evaluados en los ítems de “dominio contenidos”, “actualización y relación con realidad actual” y “presentar variados enfoques”,

De esta manera las mujeres son mejor evaluadas en ítems relacionados a la responsabilidad y la empatía, mientras que los hombres sobresalen en ítems relacionados al dominio de contenidos y conocimientos.

Tabla 7: Género del profesor²

	Puntaje Mayor	D Cohen	R Pearson	t	gl	Sig.
Puntualidad inicio - término clases	Profesor (M)	-0.14	-0.07	-6.874	8562	.000
Puntualidad entrega evaluaciones	Profesor (M)	- 0.16	-0.08	-7.950	7.090.793	.000
Asistencia regular a clases	Profesor (M)	-0.07	-0.03	-3.518	8517	.000
Dominio contenidos	Profesor (H)	0.24	0.12	10.896	5.896.897	.000
Comunicación clara	Profesor (M)	-0.01	-0.01	-.618	8556	.537
Actualización y relación con realidad actual	Profesor (H)	0.06	0.03	3.042	6.204.272	.002
Presentar variados enfoques	Profesor (H)	0.16	0.08	6.905	5.804.223	.000
Estimula participación		0	0	-.200	6.577.024	.841
Compromiso con aprendizaje	Profesor (M)	-0.07	-0.03	-2.968	6.715.994	.003
Disposición a enseñar	Profesor (M)	-0.03	-0.01	-1.415	8460	.157
Acerca conocimiento mediante ejemplos	Profesor (H)	0.01	0.01	.288	6.386.645	.774
Trato equitativo	Profesor (M)	-0.07	-0.03	-3.500	8543	.000
Consistencia entre enseñado y exigido	Profesor (M)	-0.02	-0.01	-1.424	6.550.327	.155
Materiales útiles	Profesor (M)	-0.07	-0.03	-2.857	8337	.004
Entrega programa oportuna	Profesor (M)	-0.01	-0.01	-.151	8233	.880
Entrega pautas corrección	Profesor (M)	-0.21	-0.10	-9.684	6.735.408	.000
Factor 1 Desempeño	Profesor (M)	-0.01	-0.00	-.295	5.341.702	.768
Factor 2 Responsabilidad	Profesor (M)	-0.18	-0.09	-7.925	7888	.000

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics y de Ncalculators: (Utilizado el 9 de junio del 2012 en: <http://es.ncalculators.com/>)

Respecto a los factores, el factor de “responsabilidad” como era de esperar fue significativamente diferente ya que agrupa variables donde obtienen mejor puntaje las

² Para Profesor hombre: Profesor (H); (con verde)
Para Profesor mujer: Profesor (M)

profesoras mujeres. Por otro lado, no se ven diferencias estadísticamente significativas en el factor de “Desempeño”, lo que nos muestra que ese factor abarca mucho más que sólo los ítems relacionados al conocimiento, incluyendo ítems relacionados a la empatía, por lo las tendencias hacia hombres y mujeres tienden a neutralizarse.

Respecto a los tamaños del efecto, vemos que las variables “Dominio de contenidos” y “Entrega pautas de corrección” tienen tamaños de efecto bastante considerables. En la primera variable la media más alta fue obtenida por profesores hombres, mientras que en la segunda por profesoras mujeres.

Tabla 8: Cinco variables con mayor tamaño del efecto

	D Cohen	R Pearson
Dominio contenidos	0.24	0.12
Entrega pautas corrección	-0.21	-0.10
Factor 2 Responsabilidad	-0.18	-0.09
Puntualidad entrega evaluaciones	-0.16	-0.08
Puntualidad inicio - término clases	-0.14	-0.07

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics y de Ncalculators: (Utilizado el 9 de junio del 2012 en: <http://es.ncalculators.com/>)

c. Ciclo de Estudios

En la variable “ciclo de estudios”, las diferencias entre grupos son más escasas. Sólo en nueve ítems resultó que había diferencias estadísticamente significativas entre los grupos. Los ítems con diferencias estadísticamente significativas serán “Puntualidad inicio - término clases”, “Puntualidad entrega evaluaciones”, “Dominio contenidos”, “Estimula participación”, “Compromiso con aprendizaje”, “Disposición a enseñar”, “Entrega pautas corrección”, “Factor 1 Desempeño”, “Factor 2 Responsabilidad”.

En todos los ítems que resultaron ser estadísticamente significativos los puntajes mayores fueron asignados por los estudiantes de tercero, cuarto y quinto. Excepto en el ítem “dominio de contenidos” cuyo puntaje mayor fue otorgado por estudiantes de primero y segundo.

Estos resultados nos dicen poco por sí mismos y deberán ser analizados con más detalles en las interacciones, aunque indican que los alumnos de cursos superiores tienen una mejor valoración de sus profesores que los de cursos inferiores.

Tabla 9: Ciclo de estudios

	Puntaje mayor	D Cohen	R Pearson	t	gl	Sig. (bilateral)
Puntualidad inicio - término clases	Tercero-Quinto	-0.03	-0.01	-2.099	8232	.036
Puntualidad entrega evaluaciones	Tercero-Quinto	-0.09	-0.04	-4.581	8049	.000
Asistencia regular a clases	Tercero-Quinto	-0.03	-0.01	-1.963	8186	.050
Dominio contenidos	Primero-Segundo	0.18	0.09	8.409	8232	.000
Comunicación clara	Tercero-Quinto	-0.02	-0.01	-1.221	8.118.931	.222
Actualización y relación con realidad actual	Primero-Segundo	0.02	0.01	1.315	8098	.188
Presentar variados enfoques	Primero-Segundo	0.03	0.01	1.131	7731	.258
Estimula participación	Tercero-Quinto	-0.14	-0.07	-6.139	8161	.000
Compromiso con aprendizaje	Tercero-Quinto	-0.09	-0.04	-4.130	8191	.000
Disposición a enseñar	Tercero-Quinto	-0.06	-0.03	-2.567	8.032.949	.010
Acerca conocimiento mediante ejemplos		0	0	.008	8184	.994
Trato equitativo	Tercero-Quinto	-0.01	-0.01	-.428	8.080.489	.669
Consistencia entre enseñado y exigido	Tercero-Quinto	-0.03	-0.01	-1.202	7.991.514	.229
Materiales útiles		0	0	.380	7.961.051	.704
Entrega programa oportuna	Primero-Segundo	0.03	0.01	1.541	7.924.688	.123
Entrega pautas corrección	Tercero-Quinto	-0.07	-0.03	-3.121	7960	.002
Factor 1 Desempeño	Tercero-Quinto	-0.08	-0.04	-3.673	7049	.000
Factor 2 Responsabilidad	Tercero-Quinto	-0.06	-0.03	-2.878	7615	.004

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics y de Ncalculators: (Utilizado el 9 de junio del 2012 en: <http://es.ncalculators.com/>)

El único ítem que tiene un tamaño del efecto relevante es “Dominio de contenidos” (Tabla N°10), el cual es justamente el único ítem cuyo puntaje mayor no fue asignado por el nivel de estudios de tercero a quinto sino por los primeros años. Nuevamente, habrá que ver esta

variable en mayor profundidad con sus interacciones para interpretar correctamente estos resultados.

Tabla 10: Cinco variables con mayor tamaño del efecto.

	D Cohen	R Pearson
Dominio contenidos	0.18	0.09
Estimula participación	-0.14	-0.07
Puntualidad entrega evaluaciones	-0.09	-0.04
Compromiso con aprendizaje	-0.09	-0.04
Factor 1 Desempeño	-0.08	-0.04

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics y de Ncalculators: (Utilizado el 9 de junio del 2012 en: <http://es.ncalculators.com/>)

III. Efectos de interacción

a. Análisis por ítems

Las tablas de resumen siguientes nos muestran las pruebas de los efectos inter-sujetos del modelo de 3 factores o variables independientes más sus efectos de interacción, las que contienen: las sumas de cuadrados, los grados de libertad (gl), las medias cuadráticas, los estadísticos F o niveles críticos, la significación asociada a cada estadístico F y el Eta cuadrado. Para ahorrar espacio y simplificar la interpretación, se incluirán en esta y las siguientes tablas solamente las interacciones y los efectos principales que resulten ser estadísticamente significativos. Tanto los efectos principales como las interacciones que mostraremos fueron corroborados con pruebas de chi-cuadrado a un 95% de nivel de confianza.

1. Puntualidad:

En la tabla N° 11 vemos que la única interacción que posee un efecto significativo sobre nuestra variable dependiente es la que se da entre sexo del profesor y nivel de estudios.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es el sexo del profesor (0,9), seguido de la interacción entre el nivel de estudios y el sexo del profesor (0,8) y en tercer lugar el sexo del alumno (0,1).

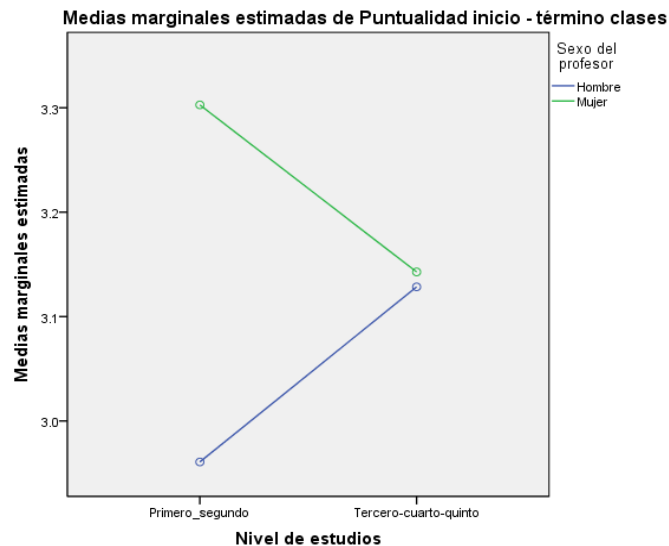
Tabla N° 11: Puntualidad inicio - término clases

Origen	Suma de cuadrados tipo III	Gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_P	43.992	1	43.992	66.989	.000	.009
SEXO_E	5.101	1	5.101	7.768	.005	.001
N_estudios * SEXO_P	37.223	1	37.223	56.681	.000	.008

a. R cuadrado = .020 (R cuadrado corregida = .019)

Respecto a la interacción entre profesor y nivel de estudios, si bien ya establecimos previamente que los profesores de sexo femenino reciben mejor evaluación en este ítem que sus pares masculinos, mediante esta interacción podemos apreciar que esas diferencias no son constantes por los años de estudio. Podemos ver que los hombres son peor evaluados los primeros años y mejor los últimos, mientras que en el caso de las mujeres ocurre todo lo contrario, es decir, son mejor evaluadas los primeros años y peor los últimos. Específicamente, los profesores hombres suben sus puntuaciones de 2.961 a 3.128 entre los primeros y los últimos años, mientras que las mujeres bajan de 3.303 a 3.143 entre los primeros y los últimos años de estudio. En el gráfico de perfiles vemos cómo el puntaje de los profesores hombres sube mientras que el de las mujeres baja.

Figura N°2: Puntualidad inicio - término clases



Ítem 2. El profesor fue puntual en la entrega de evaluaciones.

Vemos que la única interacción que posee un efecto significativo sobre nuestra variable dependiente es la que se da entre sexo del profesor y nivel de estudios. Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es el sexo del profesor (0,11), seguido de la interacción entre el nivel de estudios y el sexo del profesor (0,3) y en tercer lugar el sexo del alumno (0,1).

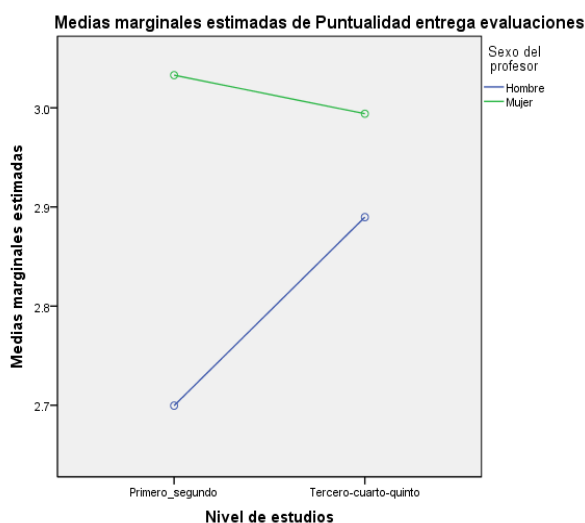
Tabla N° 12: Puntualidad entrega evaluaciones

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
N_estudios	7,824	1	7,824	10,022	,002	,001
SEXO_P	65,517	1	65,517	83,924	,000	,011
N_estudios * SEXO_P	17,941	1	17,941	22,981	,000	,003

a. R cuadrado = .019 (R cuadrado corregida = .018)

Respecto a la interacción existente entre sexo del profesor y nivel de estudios, los datos arrojaron que los hombres son peor evaluados los primeros años con una puntuación de 2.700, y mejor los últimos, con una puntuación de 2.890. Mientras que en el caso de las mujeres ocurre todo lo contrario, pasa de una puntuación de 3.033 los primeros años a 2.994 los últimos. En el gráfico de perfiles se muestra cómo los primeros años la brecha entre hombre y mujeres es bastante pronunciada y en los últimos es mucho menor.

FIGURA N° 3: Puntualidad entrega evaluaciones



Ítem 3.- El profesor asistió regularmente a clases.

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor” y “nivel de estudios”; “sexo del alumno” y “nivel de estudios” y la interacciones de los tres factores de manera simultánea.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción entre “sexo del profesor” y “nivel de estudios” (0,06), seguido de las variables “sexo del profesor” y “sexo del alumno” (0,2).

Tabla N° 13: Asistencia regular a clases

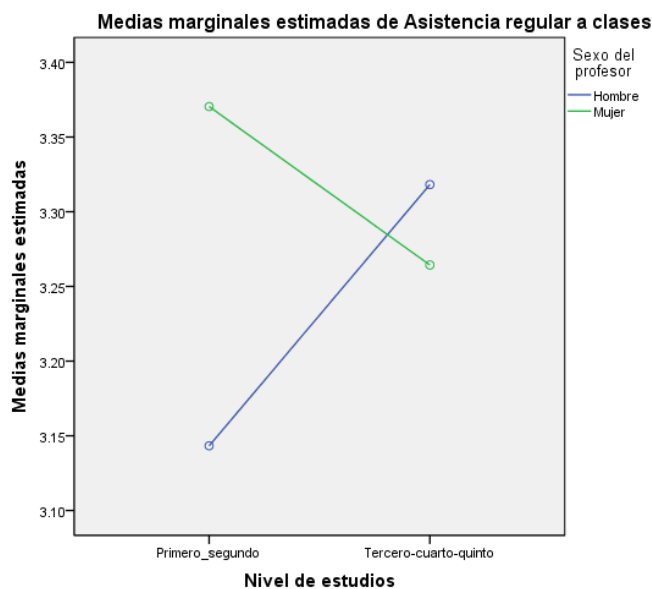
Origen	Suma de cuadrados tipo III	Gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_P	10,334	1	10,334	18,327	,000	,002
SEXO_E	9,357	1	9,357	16,595	,000	,002
N_estudios * SEXO_P	27,163	1	27,163	48,171	,000	,006
N_estudios * SEXO_E	3,382	1	3,382	5,997	,014	,001
N_estudios * SEXO_P * SEXO_E	3,442	1	3,442	6,105	,014	,001

a. R cuadrado = .015 (R cuadrado corregida = .014)

Respecto a la interacción entre sexo del profesor y nivel de estudios, podemos ver que los hombres son peor evaluados los primeros años con un puntaje de 3.143 y mejor los últimos con un puntaje de 3.318, mientras que en el caso de las mujeres ocurre todo lo contrario, obtienen un puntaje de 3.370 los primeros años y de 3.264 los últimos.

En el gráfico de perfiles se muestra cómo a los profesores de sexo femenino se les evalúa considerablemente mejor en este ítem respecto a los hombres, situación que se invierte en los últimos años. El puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja abruptamente.

FIGURA N° 4: Asistencia regular a clases



Respecto a la interacción entre Sexo del alumno y Nivel de estudios así como la triple interacción entre “Sexo del alumno”, “Sexo del profesor” y “Nivel de estudios”, encontramos que las diferencias no son constantes al combinar estos niveles. Esto quedará mejor reflejado en la tabla N° 14 que incluye la triple interacción.

Específicamente encontramos que el puntaje más bajo de los primeros años de estudio corresponde los profesores hombres que son evaluados mal por los alumnos hombres, mientras que los últimos años los peor evaluados son las profesoras mujeres justamente por los alumnos hombres.

Por otro lado, encontramos que los primeros años las mejor evaluadas son las profesoras mujeres por parte de las alumnas mujeres, mientras que en los últimos años los mejor evaluados son los profesores hombres por parte de las alumnas mujeres. Es decir, tanto alumnos hombres como alumnas mujeres invierten su lógica de evaluación entre los primeros y los últimos años de estudio.

Tabla N° 14: Nivel de estudios * Sexo del profesor * Sexo del estudiante que evalúa³

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	Masculino	3,043	0,026	2,992	3,094
		Femenino	3,244	0,019	3,207	3,28
	Mujer	Masculino	3,339	0,043	3,255	3,423
		Femenino	3,402	0,029	3,346	3,458
Tercero-cuarto-quinto	Hombre	Masculino	3,317	0,025	3,268	3,366
		Femenino	3,319	0,019	3,281	3,357
	Mujer	Masculino	3,232	0,037	3,159	3,305
		Femenino	3,296	0,022	3,252	3,34

Ítem 4.- El profesor domina los contenidos de su clase

Al ver los efectos de interacción, podemos constatar todas las interacciones posibles poseen un efecto significativo sobre nuestra variable dependiente. Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es el “sexo del profesor” (0,17), seguido de las variables “nivel de estudios (0,07) y “sexo del alumno” (0,05).

Tabla N° 15: El profesor domina los contenidos de su clase

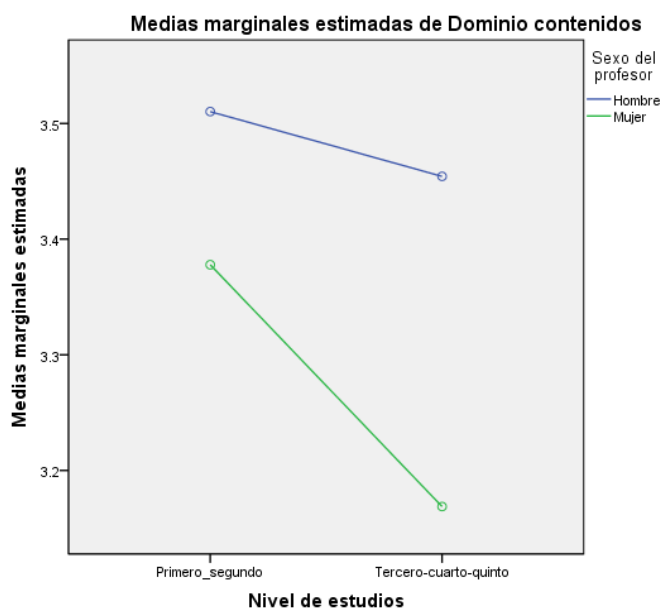
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
N_estudios	24,249	1	24,249	53,369	,000	,007
SEXO_P	60,264	1	60,264	132,635	,000	,017
SEXO_E	15,348	1	15,348	33,778	,000	,005
N_estudios * SEXO_P	8,090	1	8,090	17,805	,000	,002
N_estudios * SEXO_E	2,240	1	2,240	4,931	,026	,001
SEXO_P * SEXO_E	2,146	1	2,146	4,722	,030	,001
N_estudios * SEXO_P * SEXO_E	3,520	1	3,520	7,748	,005	,001

a. R cuadrado = .032 (R cuadrado corregida = .031)

³ En achurado los peores y mejores puntajes según cada categoría de nivel de estudios.

Respecto a la interacción entre sexo del profesor y nivel de estudios, los datos arrojaron que ambos géneros bajan su puntaje los últimos años, pero la caída es muy leve en el caso de los profesores hombres, de 3.510 a 3.454 entre los primeros y los últimos años, y un poco más abrupta en el caso de las profesoras mujeres, de 3.378 a 3.169 entre los primeros y los últimos años.

FIGURA N° 5: El profesor domina los contenidos de su clase



Respecto a la interacción entre sexo del alumno y nivel de estudios, podemos decir que la evaluación de los hombres baja de manera más abrupta que el de las mujeres (Tabla N°16). Parecido ocurre con la interacción entre sexo del alumno y sexo del profesor, donde si bien los hombres evalúan peor que las mujeres en ambas categorías, la brecha se acorta o ensancha según el sexo del profesor (Tabla N°16).

En la triple interacción, la tendencia es la misma, lo que cambia es la intensidad. Podemos ver que los profesores hombres son mejor evaluados los primeros años y los últimos, en ambos casos quienes mejor los evalúan son los estudiantes de sexo femenino. Las profesoras mujeres son peor evaluadas los primeros y los últimos años, en ambos casos quienes peor las evalúan son estudiantes de sexo masculino. Sin embargo la diferencia no es tan pronunciada entre el sexo de los profesores los primeros años, y se hace más amplia

los últimos, sobre todo porque el puntaje asignado por parte de los estudiantes hombres hacia las profesoras mujeres baja bruscamente los últimos años de estudio.

Tabla N° 16: Triple interacción- Dominio contenidos

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	Masculino	3,472	,023	3,427	3,518
		Femenino	3,548	,017	3,516	3,581
	Mujer	Masculino	3,351	,039	3,275	3,426
		Femenino	3,405	,026	3,355	3,455
Tercero-cuarto-quinto	Hombre	Masculino	3,426	,022	3,382	3,470
		Femenino	3,482	,017	3,448	3,516
	Mujer	Masculino	3,051	,033	2,986	3,116
		Femenino	3,287	,020	3,247	3,326

Ítem 5.- El profesor se comunica con claridad

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente “El profesor se comunica con claridad” son las que se dan entre “sexo del profesor” y “nivel de estudios”; entre “sexo del alumno” y “sexo del profesor”; y la interacción triple de los tres factores de manera simultánea. Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción entre “sexo del profesor” y “nivel de estudios” (0,07), seguido de la variable “sexo del alumno” (0,06).

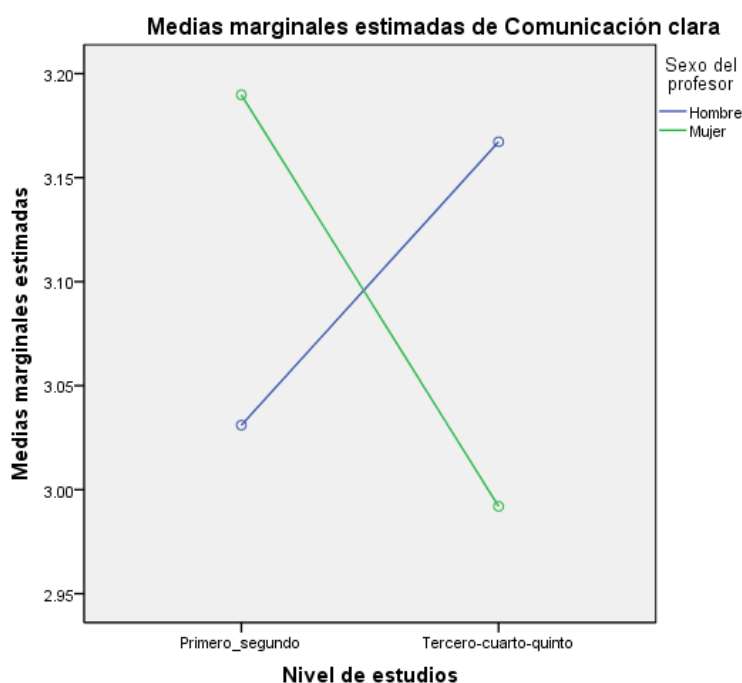
Tabla N° 17: Comunicación clara

Origen	Suma de cuadrados tipo III	Gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	33,078	1	33,078	48,182	,000	,006
N_estudios * SEXO_P	38,617	1	38,617	56,250	,000	,007
N_estudios * SEXO_E	2,698	1	2,698	3,930	,047	,001
N_estudios * SEXO_P * SEXO_E	6,882	1	6,882	10,024	,002	,001

a. R cuadrado = .015 (R cuadrado corregida = .014)

En la interacción que se da entre el sexo del profesor y nivel de estudios podemos ver que hay una relación inversa en la evaluación de profesores hombres y mujeres a lo largo del ciclo universitario. Es decir, a los hombres se les empieza evaluando más bajo, para luego mejorar, pasando de una puntuación de 3.031 a 3.167, mientras que a las mujeres se las evalúa mejor al comienzo, lo que luego empeora, pasando de 3.190 a 2.992 entre los primeros y los últimos años de estudio. El gráfico muestra cómo el puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja abruptamente.

FIGURA N° 6: El profesor se comunica con claridad



Respecto a la interacción entre sexo del alumno y nivel de estudios los datos arrojaron que las diferencias no son constantes (tabla N° 18). Los estudiantes hombres evalúan levemente peor entre un nivel y otro (3.055 a 2.980); y las mujeres levemente mejor (3.166 a 3.179). En el fondo la brecha se hace más grande en los últimos años, donde los hombres, que ya evaluaban peor que las mujeres, aumentan aún más su brecha.

Mediante la triple interacción podemos ver además que los primeros años son los profesores hombres los peor evaluados y reciben su peor evaluación por los alumnos hombres mientras que en los últimos años esa lógica se invierte y los alumnos hombres empiezan a evaluar peor a las profesoras mujeres. Paralelamente, las profesoras mujeres

son las mejores evaluadas los primeros años y reciben su mejor evaluación por parte de las alumnas mujeres, mientras que los últimos años los mejor evaluados son los profesores hombres que reciben su mejor evaluación de las alumnas mujeres.

Tabla N° 18: Triple interacción- Comunicación clara

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	Masculino	2,957	,029	2,901	3,013
		Femenino	3,105	,021	3,065	3,145
	Mujer	Masculino	3,153	,047	3,060	3,246
		Femenino	3,227	,031	3,165	3,288
Tercero-cuarto-quinto	Hombre	Masculino	3,120	,028	3,066	3,174
		Femenino	3,215	,021	3,173	3,256
	Mujer	Masculino	2,841	,041	2,761	2,920
		Femenino	3,143	,025	3,095	3,192

Ítem 6.- Los contenidos expuestos por el profesor están actualizados y tienen relación con la realidad actual.

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor” y “nivel de estudios”, y entre “sexo del alumno”, “sexo del profesor” y “nivel de estudios”; de manera simultánea.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es el “sexo del alumno” (0,07), seguido de “sexo del profesor” y de la interacción de “nivel de estudios” con “sexo del profesor” (0,02).

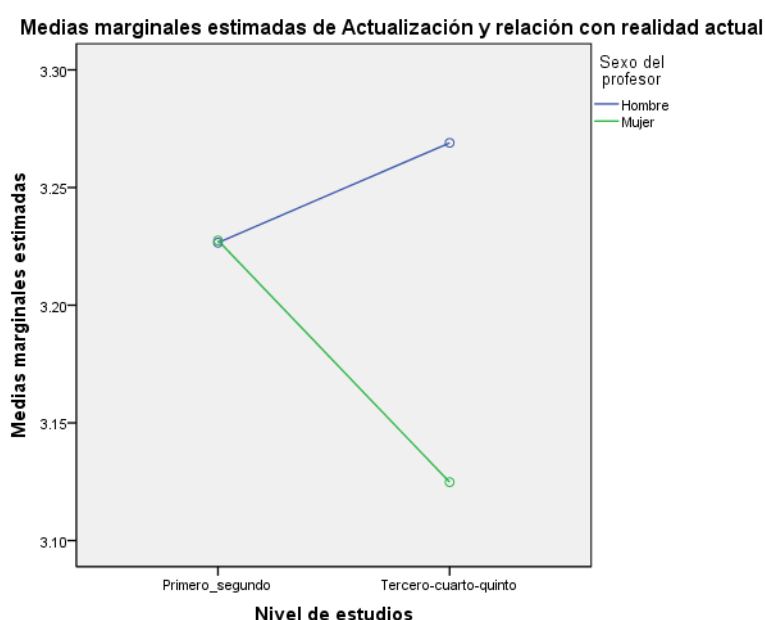
Tabla N° 19: Actualización y relación con realidad actual

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_P	6,981	1	6,981	13,696	,000	,002
SEXO_E	26,564	1	26,564	52,113	,000	,007
N_estudios * SEXO_P	7,172	1	7,172	14,069	,000	,002

a. R cuadrado = .015 (R cuadrado corregida = .014)

Respecto a la interacción entre sexo del profesor y nivel de estudios podemos ver que en los profesores el puntaje sube levemente entre los primeros y los últimos años, pasan de 3.227 los primeros años a 3.269 los últimos años, mientras que hay una leve caída el caso de las profesoras que pasan de 3.228 a 3.125 los últimos años. El gráfico de perfiles muestra más explícitamente cómo el puntaje de los profesores hombres sube levemente mientras que el de las mujeres baja levemente.

FIGURA N° 7: Actualización y relación con realidad actual



Ítem 7.- El profesor presentó variados enfoques teóricos referentes a la materia.

En la tabla N° 20, al ver los efectos de interacción, podemos constatar que todas las interacciones poseen un efecto significativo sobre nuestra variable dependiente.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es el “sexo del alumno” (0,06), seguido de la variable “sexo del profesor” (0,05).

Tabla N° 20: Presentar variados enfoques

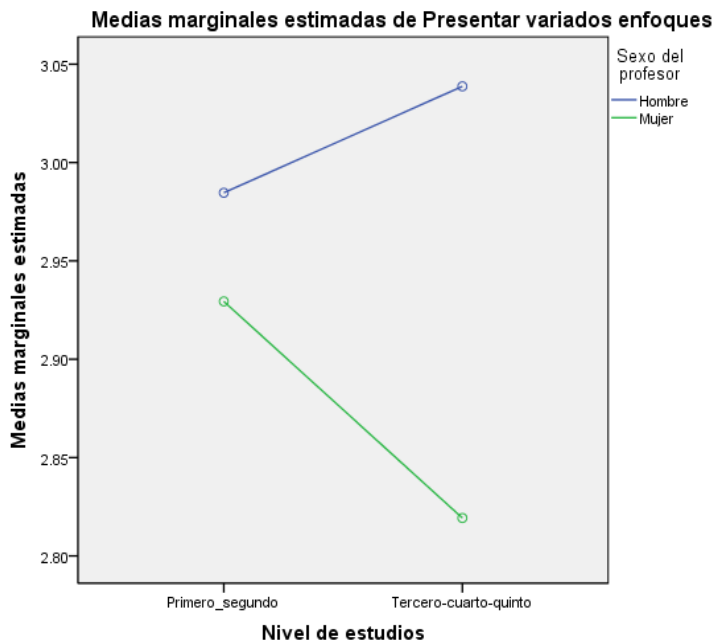
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_P	23,989	1	23,989	35,164	,000	,005
SEXO_E	29,680	1	29,680	43,506	,000	,006
N_estudios * SEXO_P	8,570	1	8,570	12,563	,000	,002
N_estudios * SEXO_E	2,612	1	2,612	3,829	,050	,001
SEXO_P * SEXO_E	2,807	1	2,807	4,115	,043	,001
N_estudios * SEXO_P * SEXO_E	2,650	1	2,650	3,884	,049	,001

a. R cuadrado = .013 (R cuadrado corregida = .012)

Respecto a la interacción que se da entre sexo del profesor y nivel de estudios, podemos ver que ambos son evaluados de manera similar los primeros años, mientras que las diferencias se dan los últimos años donde a los hombres se les evalúa levemente mejor (pasan de 2.985 los primeros años a 3.039 los últimos años) mientras que a las mujeres levemente peor (de 2.929 los primeros años a 2.819 los últimos años).

En el gráfico de perfiles vemos que el puntaje de los profesores hombres sube levemente mientras que el de las mujeres baja.

FIGURA N° 8: Presentar variados enfoques



Todas las otras interacciones quedan mejor representadas con la triple interacción (Tabla N°21), entre sexo del alumno, sexo del profesor y nivel de estudios. Además podemos ver que los profesores hombres son evaluados de manera similar a lo largo de los años, mientras que el cambio importante se da en las profesoras mujeres que son bastante peor evaluadas los últimos años, específicamente por los estudiantes hombres.

Tabla N° 21: Triple interacción- Presentar variados enfoques

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	masculino	2,932	,029	2,874	2,989
		Femenino	3,038	,021	2,996	3,079
	Mujer	masculino	2,875	,049	2,778	2,972
		Femenino	2,984	,033	2,919	3,049
Tercero-cuarto-quinto	Hombre	masculino	2,986	,028	2,931	3,041
		Femenino	3,091	,022	3,049	3,134
	Mujer	masculino	2,674	,043	2,590	2,758
		Femenino	2,965	,025	2,915	3,014

Ítem 8.- Estrategias de enseñanza que estimulan la participación

En la tabla N° 22, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”; de manera simultánea, y la interacción dada por el “nivel de estudios” y “sexo del profesor”.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,05), seguido de la variable “sexo del alumno” (0,04).

Tabla N° 22: Estrategias de enseñanza que estimulan la participación

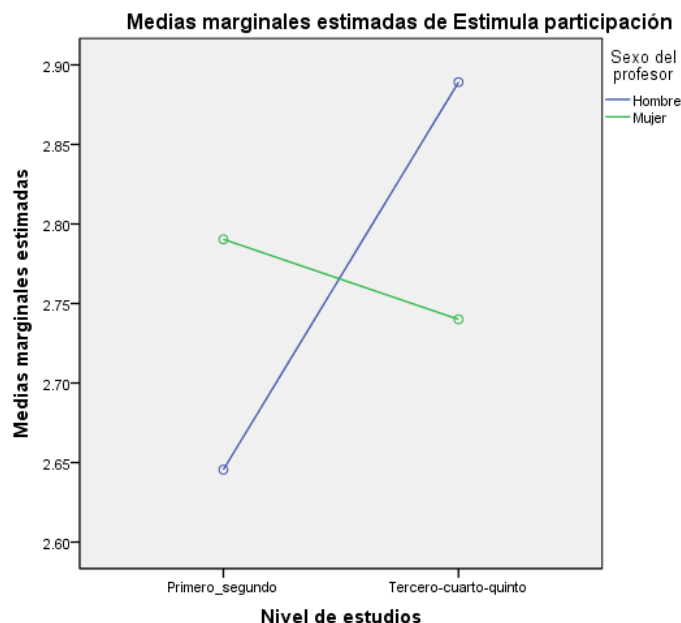
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
N_estudios	12,832	1	12,832	15,585	,000	,002
SEXO_E	26,183	1	26,183	31,802	,000	,004
N_estudios * SEXO_P	29,649	1	29,649	36,012	,000	,005
N_estudios * SEXO_P * SEXO_E	7,329	1	7,329	8,901	,003	,001

a. R cuadrado = .017 (R cuadrado corregida = .016)

Respecto a la interacción entre sexo del profesor y nivel de estudios, podemos ver que los profesores hombres son peor evaluados en este ítem los primeros años, situación que se invierte en los últimos. Se pasa de un puntaje de 2.646 a 2.889 en el caso de los hombres y de 2.790 a 2.740 en el caso de las mujeres.

El gráfico de perfiles nos muestra que en los primeros años a los profesores de sexo femenino se les evalúa mejor en este ítem, situación que se invierte en los últimos años. En definitiva, el puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja.

FIGURA N° 9: Estrategias de enseñanza que estimulan la participación



En la interacción triple (tabla N°23), podemos ver que los estudiantes hombres evalúan mal sobre todo a los profesores hombres en los primeros años, mientras que los últimos años se invierte esta lógica y empiezan a evaluar peor sobre todo a las profesoras mujeres. Por otro lado, estudiantes mujeres, que asignan los puntajes más altos en ambas categorías de nivel de estudios, los primeros años evalúan bien sobre todo a las profesoras mujeres, mientras que los últimos años comienzan a evaluar mejor a los profesores hombres.

Tabla N° 23: Interacción triple- Estimula participación

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	Masculino	2,551	,031	2,489	2,612
		Femenino	2,740	,023	2,696	2,785
	Mujer	Masculino	2,782	,052	2,681	2,884
		Femenino	2,798	,035	2,730	2,866
Tercero-cuarto-quinto	Hombre	Masculino	2,832	,030	2,773	2,892
		Femenino	2,946	,023	2,901	2,992
	Mujer	Masculino	2,624	,045	2,535	2,712
		Femenino	2,856	,027	2,803	2,909

Ítem 9.- El profesor demuestra compromiso con el aprendizaje.

En la tabla N°24, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor” con “nivel de estudios”; y la interacción entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”; de manera simultánea.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,08), seguido de la variable “sexo del alumno” (0,06)

Tabla N° 24: Compromiso con aprendizaje

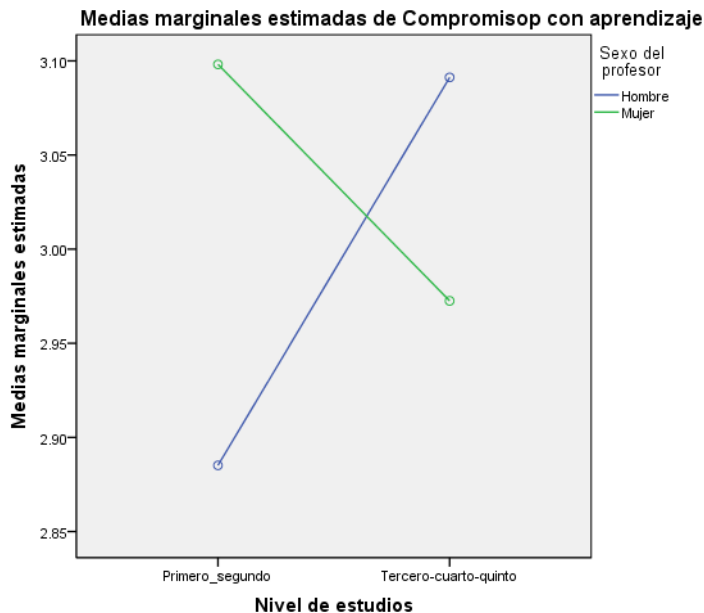
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_P	3,053	1	3,053	4,538	,033	,001
SEXO_E	30,891	1	30,891	45,927	,000	,006
N_estudios * SEXO_P	37,773	1	37,773	56,158	,000	,008
N_estudios * SEXO_P * SEXO_E	4,621	1	4,621	6,870	,009	,001

a. R cuadrado = .019 (R cuadrado corregida = .018)

Respecto a la interacción entre sexo del profesor y nivel de estudios, los hombres suben de 2.885 a 3.091 entre los primeros y los últimos años de estudio, mientras que las mujeres bajan de 3.098 a 2.973.

En el gráfico de perfiles podemos constatar cómo en los primeros años a los profesores de sexo femenino se les evalúa mejor que a los hombres, situación que cambia en los últimos años de vida universitaria del estudiante. El puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja abruptamente.

FIGURA N° 10: Compromiso con aprendizaje



Respecto a la triple interacción, podemos ver en la tabla N°25 que los estudiantes hombres evalúan mal sobre todo a los profesores hombres en los primeros años, mientras que los últimos años se invierte esta lógica y empiezan a evaluar peor sobre todo a las profesoras mujeres. Por otro lado, estudiantes mujeres, que asignan los puntajes más altos en ambas categorías de nivel de estudios, los primeros años evalúan bien sobre todo a las profesoras mujeres, mientras que los últimos años comienzan a evaluar mejor a los profesores hombres.

Tabla N° 25: Triple interacción - Compromiso con aprendizaje

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	masculino	2,787	,028	2,731	2,842
		femenino	2,983	,020	2,943	3,023
	Mujer	masculino	3,056	,047	2,964	3,148
		femenino	3,140	,031	3,079	3,202
Tercero-cuarto-quinto	Hombre	masculino	3,041	,027	2,988	3,095
		femenino	3,141	,021	3,100	3,182
	Mujer	masculino	2,863	,041	2,783	2,942
		femenino	3,082	,024	3,034	3,130

Ítem 10.- El profesor muestra disposición a enseñar y aclarar dudas.

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor” con “nivel de estudios”; y la interacción entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”; de manera simultánea.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,08), seguido de la variable “sexo del alumno” (0,06).

Tabla N° 26: Disposición a enseñar

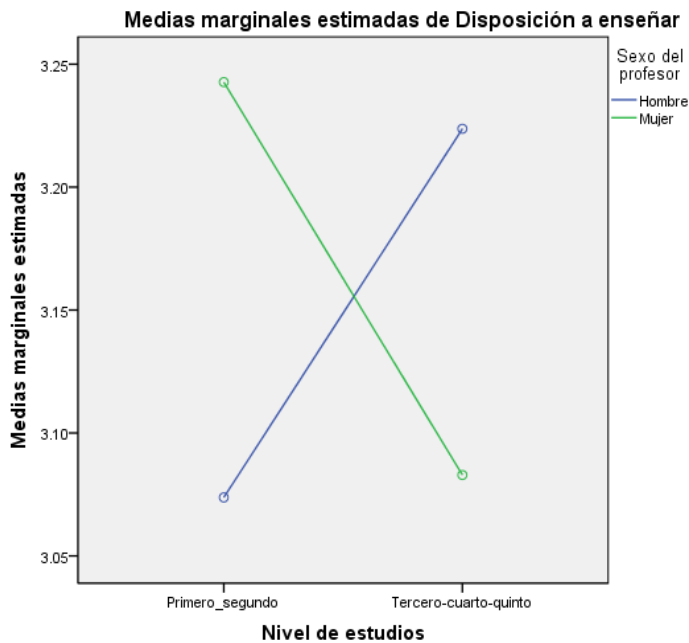
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	25,985	1	25,985	44,536	,000	,006
N_estudios * SEXO_P	32,745	1	32,745	56,122	,000	,008
N_estudios * SEXO_P * SEXO_E	3,517	1	3,517	6,028	,014	,001

a. R cuadrado = .016 (R cuadrado corregida = .015)

Respecto a la interacción entre sexo del profesor y nivel de estudios los datos arrojaron que los profesores hombres pasan de 3.074 los primeros años a 3.224 los últimos años, mientras que las mujeres bajan de 3.243 los primeros años a 3.083 los últimos.

En el gráfico de perfiles se ve cómo en los primeros años a los profesores de sexo femenino se les evalúa considerablemente mejor en este ítem respecto a los hombres, situación que se invierte en los últimos años. El puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja.

FIGURA N° 11: Disposición a enseñar



Como es ya la tendencia, en la triple interacción tabla N° 27, vemos nuevamente que los estudiantes hombres evalúan mal sobre todo a los profesores hombres en los primeros años, mientras que los últimos años se invierte esta lógica y empiezan a evaluar peor sobre todo a las profesoras mujeres. Por otro lado, las estudiantes mujeres, que asignan los puntajes más altos en ambas categorías de nivel de estudios, en los primeros años evalúan bien sobre todo a las profesoras mujeres, mientras que los últimos años comienzan a evaluar mejor a los profesores hombres.

Tabla N° 27: Triple interacción- Disposición a enseñar

Nivel de estudios	Sexo del profesor	Sexo del estudiante que evalúa	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero_segundo	Hombre	masculino	3,000	,026	2,948	3,052
		femenino	3,148	,019	3,110	3,185
	Mujer	masculino	3,209	,044	3,122	3,295
		femenino	3,277	,029	3,220	3,334
Tercero-cuarto-quinto	Hombre	masculino	3,171	,026	3,120	3,221
		femenino	3,277	,020	3,239	3,315
	Mujer	masculino	2,968	,038	2,894	3,042
		femenino	3,198	,023	3,153	3,243

Ítem 11.- El profesor acerca al estudiante al conocimiento de la disciplina

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor” con “nivel de estudios”; y la interacción entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”; de manera simultánea.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es el “sexo del alumno” (0,07), seguido de la interacción de “sexo del profesor” con “nivel de estudios” (0,04).

Tabla N° 28: Acerca conocimiento mediante ejemplos

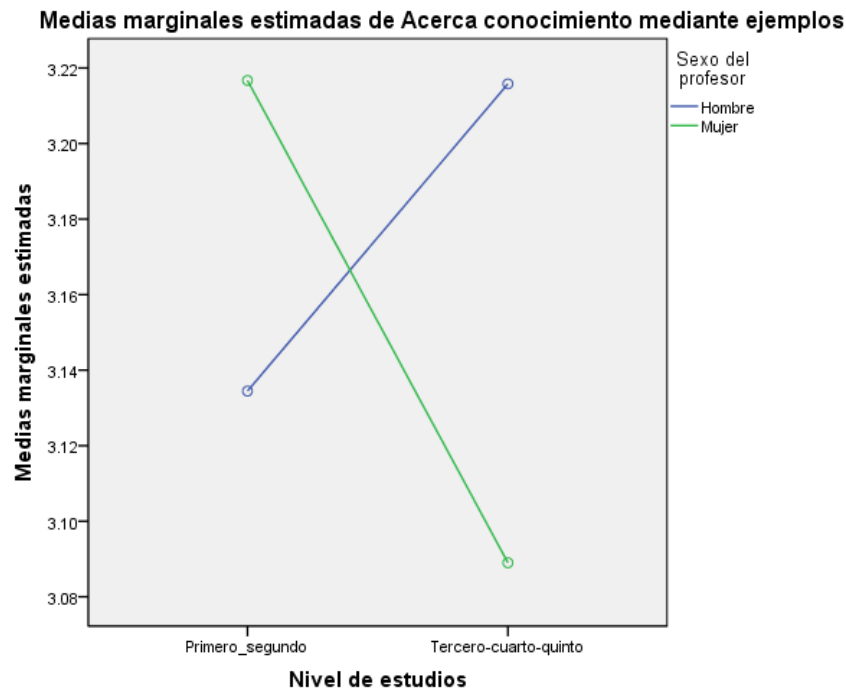
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	25,147	1	25,147	48,492	,000	,007
SEXO_P * N_estudios	15,019	1	15,019	28,962	,000	,004
SEXO_E * SEXO_P * N_estudios	2,697	1	2,697	5,200	,023	,001

a. R cuadrado = .012 (R cuadrado corregida = .011)

Respecto a la interacción entre sexo del profesor y nivel de estudios, los datos arrojaron que nuevamente los hombres son peor evaluados los primeros años, inversamente a lo que ocurre con las profesoras mujeres. Los hombres suben de 3.134 a 3.216, mientras que las mujeres bajan de 3.217 a 3.089 entre los primeros y los últimos años de estudio.

En el gráfico de perfiles vemos cómo el puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja abruptamente.

FIGURA N° 12: Acerca conocimiento mediante ejemplos



En la triple interacción tabla N° 29, nuevamente vemos que los estudiantes hombres evalúan mal sobre todo a los profesores hombres en los primeros años, mientras que los últimos años se invierte esta lógica y empiezan a evaluar peor sobre todo a las profesoras mujeres. Por otro lado, las estudiantes mujeres, que asignan los puntajes más altos en ambas categorías de nivel de estudios, los primeros años evalúan bien sobre todo a las profesoras mujeres, mientras que los últimos años comienzan a evaluar mejor a los profesores hombres.

Tabla N° 29: Triple interacción-Acerca conocimiento mediante ejemplos

Nivel de estudios	Sexo del profesor	Sexo del estudiante	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero y Segundo	Hombre	Masculino	3,054	0,025	3,005	3,102
		femenino	3,215	0,018	3,18	3,25
	Mujer	Masculino	3,174	0,041	3,093	3,255
		Femenino	3,259	0,027	3,205	3,313
Tercero cuarto y quinto	Hombre	Masculino	3,167	0,024	3,12	3,214
		Femenino	3,264	0,018	3,228	3,301
	Mujer	Masculino	2,99	0,035	2,921	3,06
		Femenino	3,188	0,021	3,146	3,23

Ítem 12.- El profesor logra un trato equitativo y respetuoso

Al ver los efectos de interacción, podemos constatar que la interacción que posee un efecto significativo sobre nuestra variable dependiente es la que se da entre “sexo del profesor” y “nivel de estudios”. Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,07), seguido de la variable “sexo del alumno” (0,05).

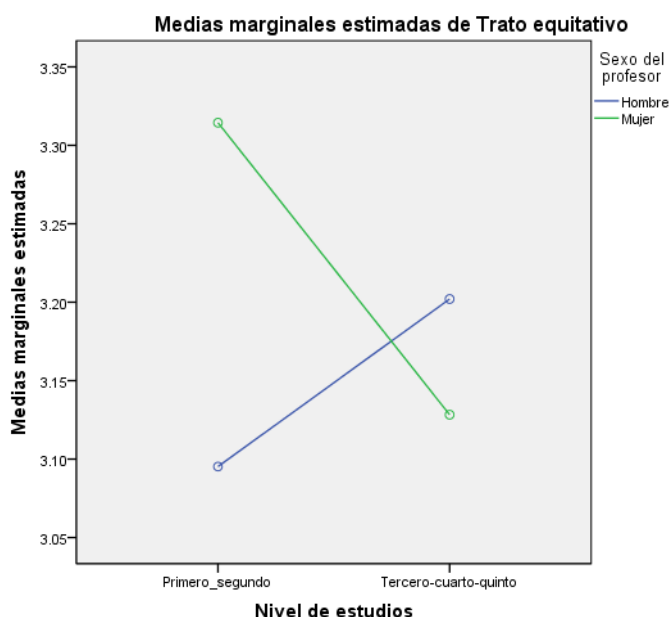
Tabla N° 30:Trato equitativo

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	21,466	1	21,466	37,004	,000	,005
SEXO_P	7,299	1	7,299	12,583	,000	,002
SEXO_P * N_estudios	29,623	1	29,623	51,065	,000	,007

a. R cuadrado = .015 (R cuadrado corregida = .014)

Respecto a la interacción entre sexo del profesor y nivel de estudios, los datos muestran que en el caso de los hombres se sube de 3.095 en los primeros años a 3.314 en los últimos años, mientras que las mujeres bajan de 3.202 los primeros años a 3.128 los últimos años. El gráfico de perfiles nos muestra más explícitamente esta relación.

FIGURA N° 13: Trato equitativo



Ítem 13.- El profesor es consistente entre lo enseñado y lo exigido

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor” con “nivel de estudios”, y la interacción entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”; de manera simultánea.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,04), seguido de la variable “sexo del alumno” (0,01).

Tabla N° 31: Consistencia entre enseñado y exigido

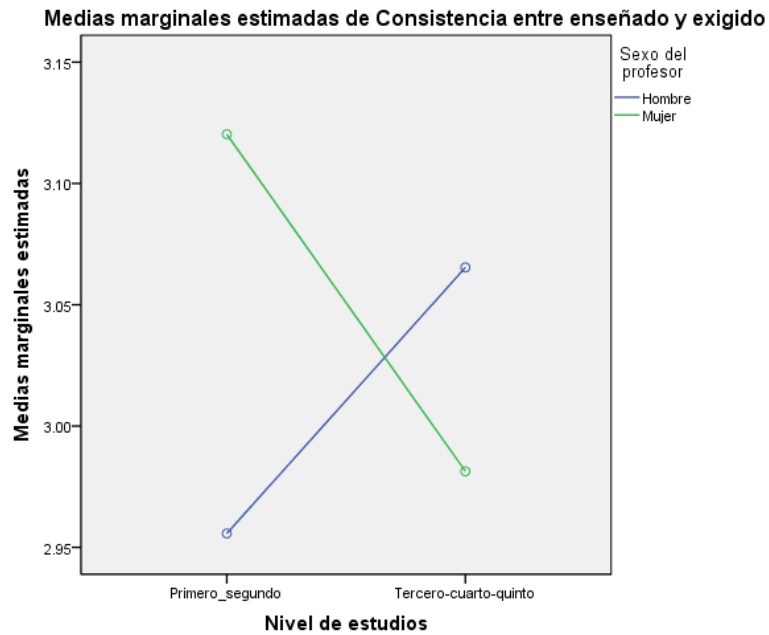
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	5,172	1	5,172	7,625	,006	,001
SEXO_P * N_estudios	21,092	1	21,092	31,096	,000	,004
SEXO_E * SEXO_P * N_estudios	3,596	1	3,596	5,301	,021	,001

a. R cuadrado = .006 (R cuadrado corregida = .005)

Respecto a la interacción que se da entre sexo del profesor y nivel de estudios, los datos arrojaron que nuevamente los hombres son peor evaluados los primeros años, inversamente a lo que ocurre con las profesoras mujeres. Los hombres suben de 2.956 a 3.120 entre los primeros y los últimos años de estudio, mientras que las mujeres bajan de 3.120 a 2.981 entre los primeros y los últimos años de estudio.

El gráfico de perfiles muestra más explícitamente cómo el puntaje de los profesores hombres sube bastante mientras que el de las mujeres cae abruptamente.

FIGURA N° 14: Consistencia entre enseñado y exigido



En la triple interacción tabla N° 32, nuevamente vemos que los estudiantes hombres evalúan mal sobre todo a los profesores hombres en los primeros años, mientras que los últimos años se invierte esta lógica y empiezan a evaluar peor sobre todo a las profesoras mujeres. Por otro lado, las estudiantes mujeres, que asignan los puntajes más altos en ambas categorías de nivel de estudios, los primeros años evalúan bien sobre todo a las profesoras mujeres, mientras que los últimos años comienzan a evaluar mejor a los profesores hombres.

Tabla N° 32: Triple Interacción- Consistencia entre enseñado y exigido

Nivel de estudios	Sexo del profesor	Sexo del estudiante	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero y segundo	Hombre	masculino	2,91	0,029	2,854	2,966
		Femenino	3,001	0,021	2,961	3,042
	Mujer	masculino	3,125	0,047	3,032	3,217
		femenino	3,116	0,032	3,054	3,178
Tercero cuarto y quinto	Hombre	masculino	3,051	0,028	2,997	3,105
		Femenino	3,08	0,021	3,038	3,122
	Mujer	masculino	2,914	0,041	2,834	2,994
		Femenino	3,049	0,025	3	3,097

Ítem 14.- El profesor prepara materiales útiles y pertinentes

Podemos ver que la interacción que posee un efecto significativo sobre nuestra variable dependiente es la que se da entre “nivel de estudios” y “sexo del profesor”. Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,06), seguido de la variable “sexo del profesor” (0,03).

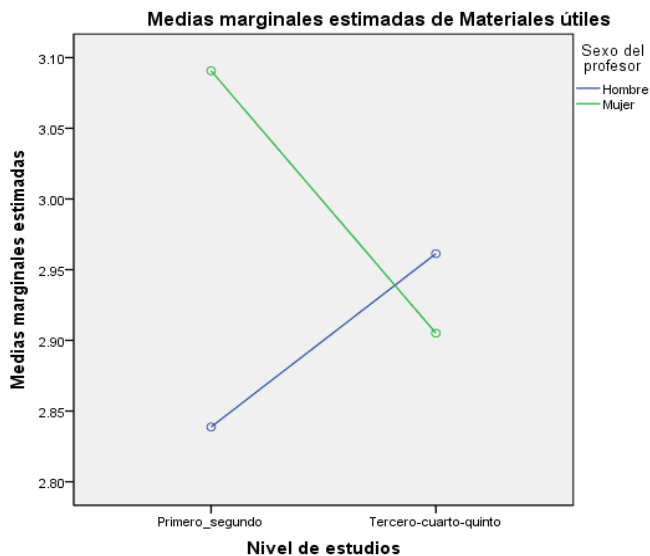
Tabla N° 33: Materiales útiles

Origen	Suma de cuadrados tipo III	Gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	10,625	1	10,625	15,539	,000	,002
SEXO_P	13,053	1	13,053	19,090	,000	,003
SEXO_P * N_estudios	32,357	1	32,357	47,320	,000	,006

a. R cuadrado = .012 (R cuadrado corregida = .011)

Respecto a la interacción entre sexo del profesor y nivel de estudios, los datos arrojaron que en el caso de los hombres se sube de 2.839 a 3.091 entre los primeros y los últimos años, mientras que las mujeres bajan de 2.961 a 2.905 entre los primeros y los últimos años de estudio. El gráfico de perfiles nos muestra más explícitamente esta relación.

FIGURA N° 15: Materiales útiles



15.- El profesor entrega el programa oportunamente

Podemos constatar que la interacción que posee un efecto significativo sobre nuestra variable dependiente es la que se da entre el “nivel de estudios” y “sexo del profesor”. Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la interacción de “sexo del profesor” con “nivel de estudios” (0,05), seguido de la variable “sexo del profesor” (0,04).

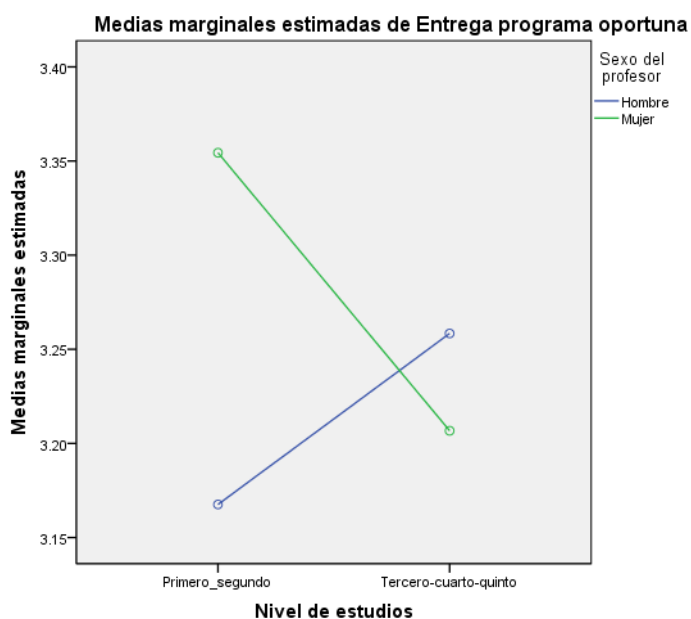
Tabla N° 34: Entrega programa oportuna

Origen	Suma de cuadrados tipo III	Gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_P	6,194	1	6,194	11,136	,001	,002
SEXO_P * N_estudios	19,300	1	19,300	34,700	,000	,005

a. R cuadrado = .007 (R cuadrado corregida = .006)

Respecto a la interacción entre el sexo del profesor y el nivel de estudios, los datos arrojaron que los profesores hombres suben de 3.168 a 3.354 entre los primeros y los últimos años, mientras que las profesoras mujeres bajan de 3.354 a 3.207 entre los primeros y últimos años de estudio. El gráfico de perfiles muestra más explícitamente esta relación.

FIGURA N° 16: Entrega programa oportuna



Ítem 16.- El profesor entregó pautas claras de corrección

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”; de manera simultánea, y la interacción dada por el “nivel de estudios” y “sexo del profesor”.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es “sexo del profesor” (0,12), en segundo lugar la interacción de “sexo del profesor” con “nivel de estudios” (0,05), seguido de la variable “sexo del alumno” (0,03).

Tabla N° 35: Entrega pautas corrección

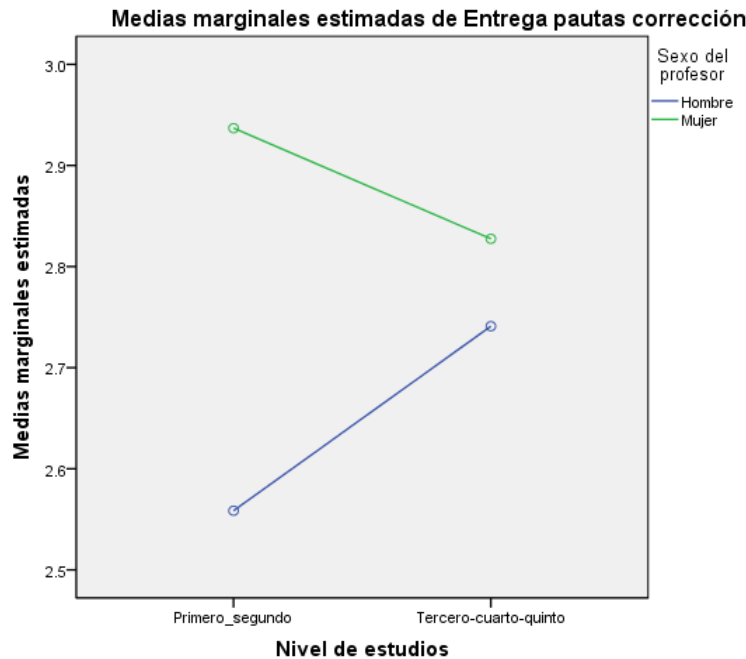
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	19,409	1	19,409	23,204	,000	,003
SEXO_P	73,168	1	73,168	87,474	,000	,012
SEXO_P * N_estudios	28,869	1	28,869	34,514	,000	,005
SEXO_E * SEXO_P * N_estudios	3,222	1	3,222	3,853	,050	,001

a. R cuadrado = .025 (R cuadrado corregida = .024)

Respecto a la interacción entre sexo del profesor y nivel de estudios, los datos arrojaron que los hombres se suben de 2.559 a 2.741 entre los primeros y los últimos años, mientras que las mujeres bajan de 2.937 a 2.827 entre los primeros y los últimos años de estudio.

El gráfico de perfiles muestra más explícitamente cómo el puntaje de los profesores hombres sube levemente mientras que el de las mujeres baja también levemente.

FIGURA N° 17: Entrega pautas corrección



En la triple interacción podemos ver que las diferencias encontradas entre ambos sexos de estudiantes no son constantes por los años de estudio así como tampoco por el sexo del

profesor. Sin embargo, a diferencia de otras interacciones más puras, en este caso la tendencia es la misma y lo que cambia es la intensidad. Lo constante es que tanto en los primeros como en los últimos años las profesoras mujeres son las mejor evaluadas por parte de las alumnas mujeres. Lo que cambia, es que los profesores hombres son evaluados especialmente mal por los alumnos hombres en los primeros años, mientras que los últimos años alumnos hombres equiparan la evaluación que dan a sus profesores hombres con respecto a la evaluación que dan a las profesoras mujeres.

Tabla N° 36: Triple interacción- Entrega pautas corrección

Nivel de estudios	Sexo del profesor	Sexo del estudiante	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero y segundo	Hombre	masculino	2,472	0,032	2,41	2,534
		femenino	2,645	0,023	2,6	2,69
	Mujer	masculino	2,886	0,052	2,783	2,988
		femenino	2,988	0,035	2,919	3,057
Tercero, cuarto y quinto	Hombre	masculino	2,721	0,031	2,66	2,782
		femenino	2,761	0,024	2,714	2,808
	Mujer	masculino	2,746	0,046	2,656	2,835
		femenino	2,909	0,028	2,855	2,963

b. Análisis agrupado por factores

Factor 1 – Desempeño en el aula

Al ver los efectos de interacción, podemos constatar que las interacciones que poseen un efecto significativo sobre nuestra variable dependiente son las que se dan entre “sexo del profesor”, “sexo del alumno” y “nivel de estudios”, de manera simultánea, y la interacción dada por el “nivel de estudios” y “sexo del profesor”.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es “sexo del alumno” (0,07), seguido de la interacción de “sexo del profesor” con “nivel de estudios” (0,03).

Tabla N° 37: Factor 1 – Desempeño en el aula

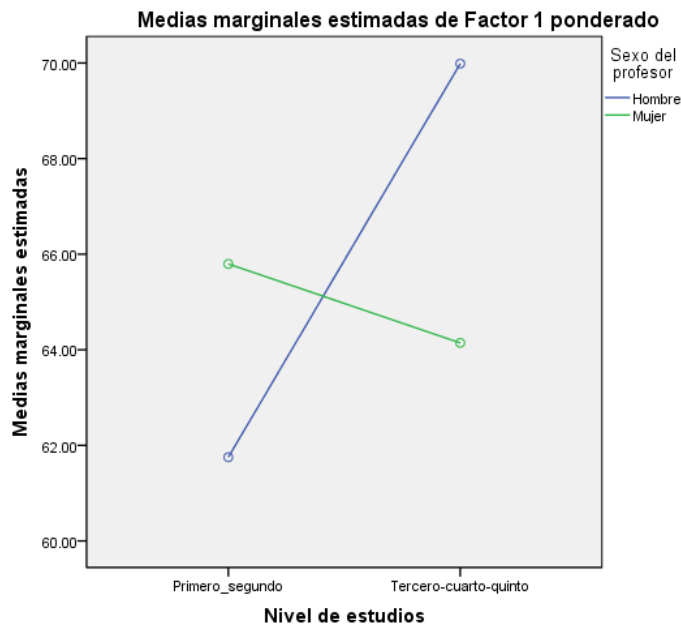
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	65268,049	1	65268,049	43,549	,000	,007
N_estudios	12780,522	1	12780,522	8,528	,004	,001
SEXO_P * N_estudios	28816,511	1	28816,511	19,227	,000	,003
SEXO_E * SEXO_P *	7323,758	1	7323,758	4,887	,027	,001
N_estudios						

a. R cuadrado = .015 (R cuadrado corregida = .014)

Respecto a la interacción entre sexo del profesor y nivel de estudios los datos arrojaron que en el caso de los hombres se sube de un puntaje de 61.752 a 69.988 entre los primeros y los últimos años, mientras que las mujeres bajan de un puntaje de 65.795 a 64.144 entre los primeros y los últimos años de estudio.⁴

En el gráfico de perfiles aparecen representadas las medias de la puntuación calculadas en cada subgrupo resultante de combinar cada nivel de la variable “sexo del profesor” con cada nivel de la variable “nivel de estudios”. Se ve que el puntaje de los profesores hombres sube bastante mientras que el de las mujeres baja.

FIGURA N° 18: Factor 1 – Desempeño en el aula



⁴ Estas medidas no representa la escala de evaluación, sino sumatorias agregadas, donde el límite inferior es de 60.033 y el superior de 71.698.

En la triple interacción tabla N° 38, vemos en el primer factor se sigue la misma tendencia que la ya observada en la mayoría de los ítems. Vale decir, los estudiantes hombres evalúan peor en ambas categorías de la variable nivel de estudios, pero en los primeros años evalúan mal sobre todo a los profesores hombres y especialmente a las profesoras mujeres en los últimos.

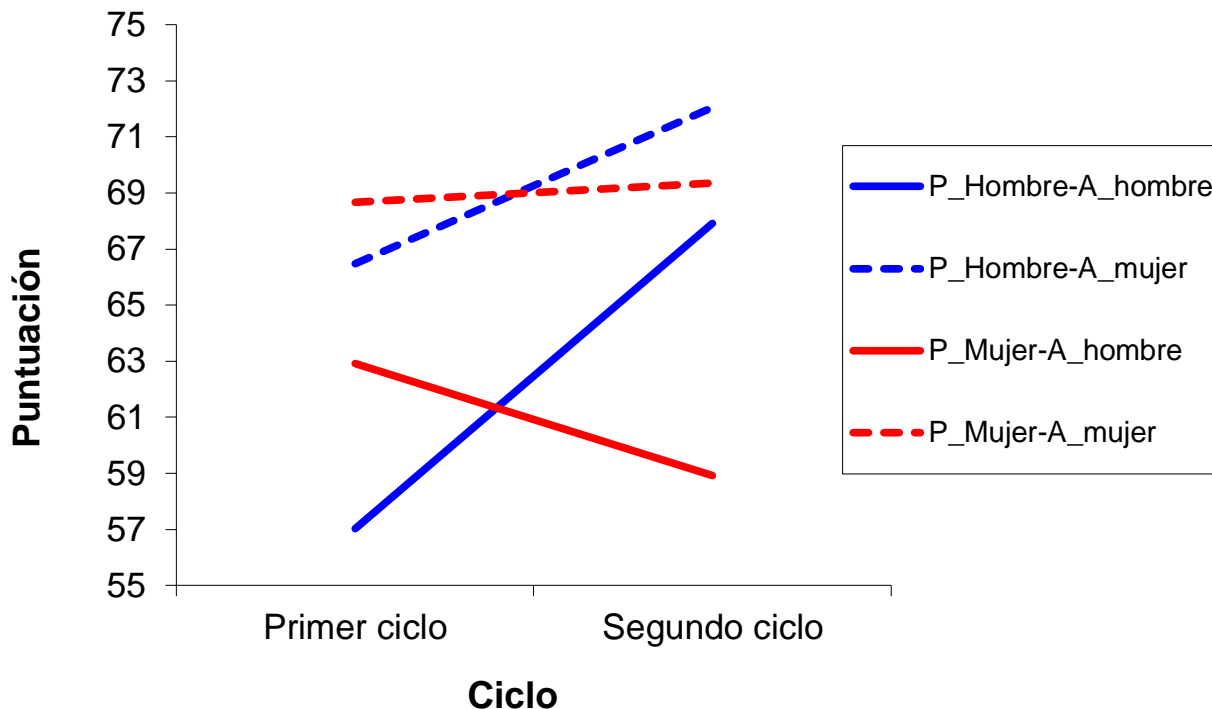
Por otro lado, las estudiantes mujeres, que asignan los puntajes más altos en ambas categorías de nivel de estudios, los primeros años evalúan bien sobre todo a las profesoras mujeres, mientras que los últimos años comienzan a evaluar mejor a los profesores hombres.

Tabla N° 38: Triple interacción-Factor 1 – Desempeño en el aula

Nivel de estudios	Sexo del profesor	Sexo del estudiante	Media	Error típ.	Intervalo de confianza 95%	
					Límite inferior	Límite superior
Primero y segundo	Hombre	masculino	57,035	1,416	54,26	59,81
		femenino	66,47	1,036	64,439	68,5
	Mujer	masculino	62,912	2,383	58,241	67,582
		femenino	68,679	1,617	65,508	71,849
Tercero, cuarto y quinto	Hombre	masculino	67,924	1,39	65,2	70,648
		femenino	72,052	1,055	69,984	74,12
	Mujer	masculino	58,929	2,103	54,807	63,051
		Femenino	69,358	1,225	66,956	71,761

El siguiente gráfico (Figura N° 19) representa lo que ha sido la tendencia de la mayoría de los ítems agrupados en este factor, vale decir, la evaluación hacia los profesores hombres y mujeres se invierte en el paso de un ciclo de estudios al otro. Esta inversión se da tanto en estudiantes masculinos como femeninos, sin embargo se mantiene la tendencia general respecto a que los estudiantes hombres evalúan peor que las mujeres tanto en el primer como en el segundo ciclo de estudios.

Figura N° 19: Triple interacción Factor 1Desempeño



Factor 2- Responsabilidad.

Al ver los efectos de interacción, podemos constatar que la única interacción que posee un efecto significativo sobre nuestra variable dependiente es la que se da entre “nivel de estudios” y “sexo del profesor”.

Respecto a los tamaños de los efectos, vemos que el factor que más varianza explica de la variable dependiente es la “sexo del profesor” (0,11), seguido de la interacción de “sexo del profesor” con “nivel de estudios” (0,08).

Tabla N° 39: Factor 2- Responsabilidad.

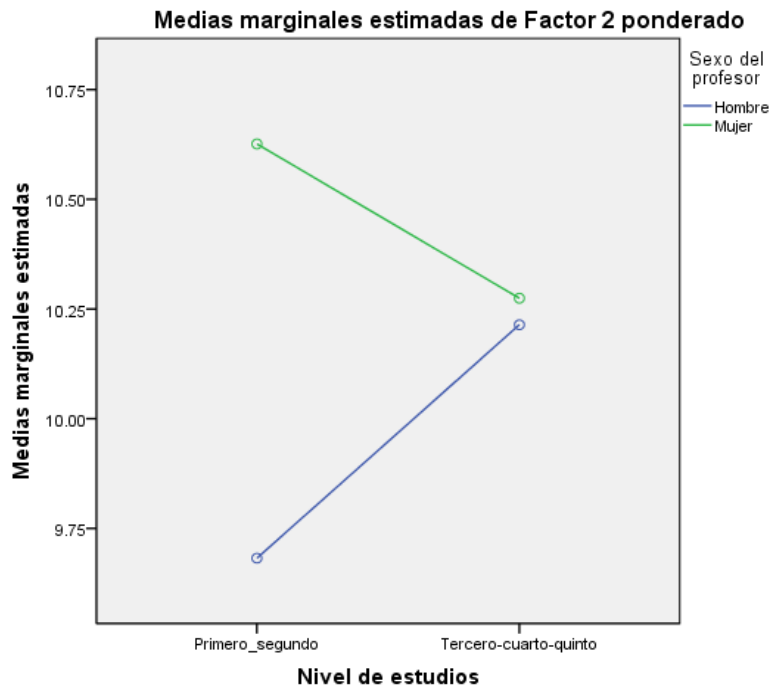
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
SEXO_E	44,096	1	44,096	10,138	,001	,001
SEXO_P	330,742	1	330,742	76,037	,000	,011
SEXO_P * N_estudios	255,958	1	255,958	58,844	,000	,008

a. R cuadrado = .025 (R cuadrado corregida = .024)

Respecto a la interacción entre sexo del profesor y nivel de estudios, los datos mostraron que los hombres suben de 9.682 a 10.214 entre los primeros y los últimos años, mientras que las mujeres bajan de 10.626 a 10.275 entre los primeros y los últimos años de estudios.⁵

En el gráfico de perfiles se muestra mejor esta relación.

FIGURA N° 20: Factor 2- Responsabilidad



⁵ Estas medidas no representan la escala de evaluación, sino sumatorias agregadas, donde el límite inferior es de 9.593 y el superior de 10.768.

IV. Síntesis de resultados

Para tener una visión general más simple de los principales resultados de los efectos de interacción, resumimos la información en las dos tablas siguientes.

Para una correcta interpretación de las columnas a continuación hay que considerar ciertos puntos:

1. La columna “interacción” incluirá los nombres de las dos o tres variables que están interactuando. A su lado aparece la columna que considera la variable de segmentación que nos permitirá comparar y evaluar la interacción en relación a esas dos categorías.
2. En la columna “puntaje mayor” se incluye la dirección de la interacción cuya nomenclatura es la misma que en las tablas anteriores (ej: profesor (H)). Por lo tanto, la categoría de la variable interactuando que se incluya en dicha columna será la que obtenga la media más alta en relación al puntaje de evaluación docente.
3. Cuando en ambas categorías de la variable de segmentación, aparezca la misma categoría de la otra variable interactuando, (Ej: profesor (M)-profesor (M)), se incluirá un signo adicional al lado de una de la categoría de más abajo (Ej: Profesor (M) (-)). Cuando el signo sea negativo, eso quiere decir que, si bien la misma categoría sigue obteniendo la media más alta, la diferencia entre ella y su categoría de referencia disminuye entre ambas categorías de segmentación. Cuando el signo es positivo, quiere decir que esas diferencias aumentan a lo largo de las categorías de segmentación.
4. Cuando hayan tres variables interactuando, nuevamente la variable de segmentación nos permitirá observar la media más alta, pero esta vez no será la media de las categorías de la variable interactuando, sino que la media de la interacción de las categorías de las otras dos variables. (Ej: Profesor (H) *Alumno (M)). Podríamos decir, por ejemplo, que entre primero y segundo el puntaje más alto se da por la interacción entre “profesor mujer*alumno mujer”, mientras que entre tercero a quinto el mayor puntaje más alto se da por la interacción entre “profesor hombre*alumno mujer”.
- 5- Nuevamente, cuando el puntaje obtenido en ambas categorías de la variable de segmentación refiera a la misma interacción de categorías de las variables, se incluirá un signo (+) o (-) que haga alusión a si la segunda interacción aumenta las diferencia de medias o la disminuye en relación a las categorías de referencia.

Tabla 40: Principales resultados de los efectos de interacción. -1

Variable	Interacción	Variable de Segmentación	Puntaje mayor	F	Sig.	Eta al cuadrado parcial	
El profesor inició y terminó puntualmente su clase.	Nivel *Profesor	Primero Segundo	Profesor (M)	56,681	.000	.008	
		Tercero-Quinto	Profesor (H)				
El profesor fue puntual en la entrega de evaluaciones.	Nivel *Profesor	Primero Segundo	Profesor (M)	22,98	0,00	0,003	
		Tercero-Quinto	Profesor (M) (-)				
El profesor asistió regularmente a clases.	Nivel *Profesor	Primero Segundo	Profesor (M)	48,17	0,00	0,006	
		Tercero-Quinto	Profesor (H)				
	Nivel*Alumno	Primero Segundo	Alumno (M)	5,997	0,01	0,001	
		Tercero-Quinto	Alumno (M) (-)				
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	6,105	0,01	0,001	
		Tercero-Quinto	Profesor (H)* Alumno (M)				
El profesor domina los contenidos de su clase.	Nivel *Profesor	Primero Segundo	Profesor (H)	17,81	0,00	0,002	
		Tercero-Quinto	Profesor (H)(+)				
	Nivel*Alumno	Primero Segundo	Alumno (M)	4,931	0,03	0,001	
		Tercero-Quinto	Alumno (M) (+)				
	Alumno*Profesor	Alumno (H)	Profesor (H)	4,722	0,03	0,001	
		Alumno (M)	Profesor (H) (-)				
	Nivel*Profesor*alumno	Primero Segundo	Profesor (H)* Alumno (M)	7,748	0,01	0,001	
		Tercero-Quinto	Profesor (H) *Alumno (M) (-)				
	El profesor se comunica con claridad	Nivel *Profesor	Primero Segundo	Profesor (M)	56,25	0,00	0,007
			Tercero-Quinto	Profesor (H)			
Nivel*Alumno		Primero Segundo	Alumno (M)	3,93	0,05	0,001	
		Tercero-Quinto	Alumno (M) (+)				
Nivel*Profesor*alumno		Primero Segundo	Profesor (M) * Alumno (M)	10,02	0,00	0,001	
		Tercero-Quinto	Profesor (H) * Alumno (M)				
Actualización de contenidos	Nivel *Profesor	Primero Segundo	Profesor (M)	14,07	0,00	0,002	
		Tercero-Quinto	Profesor (H)				
El profesor presentó variados enfoques teóricos referentes a la materia.	Nivel *Profesor	Primero Segundo	Profesor (H)	12,56	0,00	0,002	
		Tercero-Quinto	Profesor (H)(+)				
	Nivel*Alumno	Primero Segundo	Alumno (M)	3,829	0,05	0,001	
		Tercero-Quinto	Alumno (M) (+)				
	Alumno*Profesor	Alumno (H)	Profesor (H)	4,115	0,04	0,001	
		Alumno (M)	Profesor (H) (-)				
	Nivel*Profesor*alumno	Primero Segundo	Profesor (H) * Alumno (M)	3,884	0,05	0,001	
		Tercero Quinto	Profesor (H) * Alumno (M) (+)				

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics, versión 19 para Windows

Tabla 41: Principales resultados de los efectos de interacción. -2

Variable	Interacción	Variable de Segmentación	Puntaje mayor	F	Sig.	Eta al cuadrado parcial
Estrategias de enseñanza que estimulan la participación	Nivel *Profesor	Primero Segundo	Profesor (M)	36,01	0,00	0,01
		Tercero-Quinto	Profesor (H)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	8,90	0,00	0,00
		Tercero-Quinto	Profesor (H)* Alumno (M)			
El profesor demuestra compromiso con el aprendizaje.	Nivel *Profesor	Primero Segundo	Profesor (M)	56,16	0,00	0,01
		Tercero-Quinto	Profesor (H)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	6,87	0,01	0,00
		Tercero-Quinto	Profesor (H)* Alumno (M)			
El profesor muestra disposición a enseñar y aclarar dudas.	Nivel *Profesor	Primero Segundo	Profesor (M)	56,12	0,00	0,01
		Tercero-Quinto	Profesor (H)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	6,03	0,01	0,00
		Tercero-Quinto	Profesor (H)* Alumno (M)			
El profesor acerca al estudiante al conocimiento de la disciplina	Nivel *Profesor	Primero Segundo	Profesor (M)	28,96	0,00	0,00
		Tercero-Quinto	Profesor (H)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	5,20	0,02	0,00
		Tercero-Quinto	Profesor (H)* Alumno (M)			
El profesor logra un trato equitativo y respetuoso	Nivel *Profesor	Primero Segundo	Profesor (M)	51,07	0,00	0,01
		Tercero-Quinto	Profesor (H)			
El profesor es consistente entre lo enseñado y lo exigido	Nivel *Profesor	Primero Segundo	Profesor (M)	31,10	0,00	0,00
		Tercero-Quinto	Profesor (H)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (H)	5,30	0,02	0,00
		Tercero-Quinto	Profesor (H)* Alumno (M)			
El profesor prepara materiales útiles	Nivel *Profesor	Primero Segundo	Profesor (M)	47,32	0,00	0,01
		Tercero-Quinto	Profesor (H)			
El profesor entrega el programa oportunamente	Nivel *Profesor	Primero Segundo	Profesor (M)	34,70	0,00	0,01
		Tercero-Quinto	Profesor (H)			
El profesor entregó pautas claras de corrección	Nivel *Profesor	Primero Segundo	Profesor (M)	34,51	0,00	0,01
		Tercero-Quinto	Profesor (M) (-)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	3,85	0,05	0,00
		Tercero-Quinto	Profesor (M) *Alumno (M) (-)			
Factor 1 – Desempeño en el aula	Nivel *Profesor	Primero Segundo	Profesor (M)	19,23	0,00	0,00
		Tercero-Quinto	Profesor (H)			
	Nivel*Profesor*alumno	Primero Segundo	Profesor (M) * Alumno (M)	4,89	0,03	0,00
		Tercero-Quinto	Profesor (H)* Alumno (M)			
Factor 2- Responsabilidad.	Nivel *Profesor	Primero Segundo	Profesor (M)	58,84	0,00	0,01
		Tercero-Quinto	Profesor (M) (-)			

Fuente: Elaboración propia en base a las herramientas estadísticas de IBM SPSS Statistics, versión 19 para Windows

Capítulo 5: Discusión y Conclusiones

Lo primero que podemos concluir a grandes rasgos, es que nuestra estrategia de análisis nos permitió observar patrones de género tanto en evaluadores como en evaluados. Pasamos a relatar a continuación de forma más específica los puntos principales de nuestra interpretación:

En primer lugar, respecto al ciclo de estudios o nivel de estudio de la carrera, si bien los resultados de estudios internacionales dan gran preponderancia a este factor en tanto permite diferenciar claramente la importancia asignada a ciertos ítems entre alumnos de los primero y los últimos años (Hills, Naegle y Bartkus, 1997, 2009); nuestros resultados hablan de una relativa consistencia en cuanto a la direccionalidad, lo que no permite identificar diferencias sustantivas entre ciertos grupos de ítems en relación a las categorías de la variable independiente. En casi todos los ítems que resultaron ser estadísticamente significativos, los puntajes mayores fueron asignados por los estudiantes de tercero, cuarto y quinto. Solamente en el ítem “dominio de contenidos” su puntaje mayor fue otorgado por estudiantes de primero y segundo. Lo que implica que en general, los alumnos de los primeros años son más críticos con sus profesores que los alumnos de años superiores.

Si bien esta variable cobrará mayor importancia en cuanto a sus interacciones que describiremos más adelante, se puede adelantar que una interpretación posible es que el aumento en los puntajes asignados a profesores masculinos, principalmente por estudiantes mujeres los últimos años, fue lo suficientemente importante como para arrastrar a todos los puntajes en general a ser en promedio mayores los últimos años respecto a los primeros.

En cuanto al tamaño del efecto, el que resultó ser más significativo es “Dominio de contenidos”, el cual es justamente el único ítem cuyo puntaje mayor no fue asignado por el nivel de estudios de tercero a quinto sino por los primeros años.

De esta manera, aceptamos la primera hipótesis respecto a que los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile varían significativamente en relación al nivel o año cursado por el alumno.

Respecto a las hipótesis referidas al género del estudiante, los primeros resultados nos mostraron de manera bastante clara que las mujeres como estudiantes tienden a evaluar mejor que los hombres en casi todos los ítems, incluidos los dos factores. También en casi todos ellos se encontraron que las diferencias entre ambos géneros de estudiantes eran estadísticamente significativas. Esto en términos teóricos ya implica un sesgo de género. Es posible pensar en un sesgo de aquiescencia de las mujeres, que tienden a ser más condescendientes con sus profesores.

De acuerdo a la revisión bibliográfica esto no es para nada nuevo. En la literatura revisada, ya otros estudios habían dado cuenta de que las estudiantes mujeres evalúan mejor que los hombres en ciertos ítems, sobre todo luego del meta análisis de Aleamoni y Hexner (1980), donde los autores citan varios estudios realizados hasta la fecha que habían llegado a la conclusión de que las estudiantes mujeres tienden a evaluar mejor que los estudiantes hombres. En nuestro caso de estudio, aquello parece replicarse pero de forma aún más global, las mujeres evalúan significativamente mejor en prácticamente todo. Sólo en dos ítems no había diferencias significativas entre ambos géneros: “Puntualidad en la entrega de evaluaciones” y “Entrega el programa de forma oportuna” que son posiblemente los dos ítems de responsabilidad más fácilmente operacionalizables por los estudiantes en conductas concretas del profesor.

Por lo tanto se acepta la segunda hipótesis respecto a que los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, varían significativamente en relación al género del alumno, aunque la direccionalidad de esa diferencia sea siempre la misma.

Respecto al género del profesor, tomando en cuenta sólo los ítems con diferencias estadísticamente significativas, en ocho ítems y un factor las profesoras mujeres resultaron ser mejor evaluadas que los profesores hombres, estos son “Puntualidad en el inicio y termino de las clases”, “Puntualidad en la entrega de evaluaciones”, “Asistencia regular a clases”, “Compromiso con el aprendizaje”, “Disposición a enseñar”, “Trato equitativo”, “Materiales útiles”, “Entrega pautas corrección”, “Factor 2 Responsabilidad”.

Los profesores hombres por su parte fueron mejor evaluados en 3 ítems, estos son: “Dominio de contenidos”, “Actualización y relación con realidad actual” y “Presenta variados enfoques”.

No se observaron diferencias estadísticamente significativas en los otros 6 ítems ni en el factor de “Desempeño”, esto último nos muestra la diversidad de ítems que agrupa este factor, ya que si bien a los hombres se les evaluó mejor en los ítems relacionados al conocimiento, a las mujeres se les evaluó mejor en ítems asociados a la empatía y al trato con los estudiantes, por lo que se entiende que el factor se neutralice en la sumatoria final de los ítems.

En resumen, a los hombres se les evalúa mejor en los ítems más disciplinarios, mientras que las mujeres son mejor evaluadas en los ítems referentes a responsabilidad, pero también en aquellos que tienen que ver con disposición a enseñar y empatía.

Respecto a los tamaños del efecto, considerando significativos los tamaños del efecto cercanos a 0.20, o 0.10 en R de Pearson, vemos que los ítems “Dominio de contenidos” y “Entrega pautas de corrección” tienen tamaños de efecto bastante considerables (0.24 y 0.21 respectivamente). En el primer ítem el puntaje más alto fue obtenido profesores hombres, mientras que en el segundo ítem por profesoras mujeres. Resulta interesante como ambas variables, junto con tener los mayores tamaños del efecto, pueden ser bastante representativas de las formas de evaluar a cada género.

El hecho de que en términos generales a las profesoras mujeres se les evalúa mejor que a sus pares masculinos se condice con lo señalado por los autores revisados en el marco teórico (Tatro, 1995; Basow, 1998; Smith et al 2007) que señalan que en áreas como las humanidades, ciencias sociales y comunicaciones, la mayoría de los estudios han demostrado que estudiantes hombres y estudiantes mujeres tienden a evaluar a sus profesoras mujeres mejor que a sus profesores hombres en general.

Sin embargo, y coincidente con lo señalado por la socióloga Estadounidense Susan Basow (1998), más que esperar grandes tamaños de efectos estadísticos, los cuales por lo general son pequeños y suelen favorecer a las mujeres en términos generales, lo que interesa es

analizar los efectos de interacción entre las variables de género y otras variables contextuales, ya que allí, en el ámbito específico, es dónde se pueden analizar los patrones de género de forma más sustantiva. De manera tal que el núcleo de nuestro análisis estuvo puesto en los resultados de la interacción de estas tres variables: género del profesor, género del alumno y nivel de estudios.

Respecto a la interacción entre profesor y nivel de estudios, si bien ya establecimos previamente que los profesores de sexo femenino reciben mejor evaluación que sus pares masculinos, mediante esta interacción pudimos ver que esas diferencias no son constantes a lo largo de los años de estudio. Los datos mostraron que en general los hombres son peor evaluados los primeros años y mejor los últimos, mientras que en el caso de las mujeres ocurre todo lo contrario, es decir, son mejor evaluadas los primeros años y peor los últimos. De esta manera en 14 de los 16 ítems los profesores hombres son mejor evaluados que las profesoras mujeres los últimos años del ciclo de estudios, vale decir, en tercero, cuarto y quinto de universidad.

Se puede ver que incluso en ítems que tienen relación con lo que la literatura entiende como propiamente femeninos, relativos a la empatía y responsabilidad, los profesores hombres obtuvieron puntajes mayores en la evaluación de los cursos superiores.

Específicamente, los ítems en que los hombres son mejor evaluados por los estudiantes de los últimos años son: “Puntualidad en el inicio y termino de las clases”, “Asistencia regular a clases”, “Comunicación clara”, “Dominio de contenidos”, “Presenta variados enfoques” “Actualización de contenidos”, “Estimula la participación”, “Compromiso con el aprendizaje”, “Disposición a enseñar”, “Acerca conocimiento mediante ejemplos”, “Trato equitativo”, “Consistencia entre lo enseñado y lo exigido”, “Materiales útiles”, “Entrega el programa oportunamente”.

De estos ítems los únicos en los que no se cruzan las tendencias entre los primeros y los últimos años, vale decir que no se da una forma de cruz en los gráficos, son los ítems: “Dominio de contenidos” y “Presenta variados enfoques”, que son justamente los únicos en los que los profesores hombres son mejor evaluados también los primeros años, por lo que

los últimos años esa diferencia se ensancha aún más al aumentar los profesores hombres sus puntajes y al bajar las profesoras mujeres el suyo.

Los únicos ítems en los que las profesoras mujeres logran mantener puntajes mejores que los profesores hombres son sólo dos y están estrictamente relacionados con el factor de responsabilidad, y son “Puntualidad entrega evaluaciones” y “Entrega pautas corrección”. Aun así, las diferencias entre sexos se acortan en los últimos años.

Finalmente respecto a los factores, como es de esperar se pudo observar que en el factor de “Desempeño” los hombres son mejor evaluados los últimos años lo que se condice con los ítems en general. Respecto al factor de “responsabilidad”, vemos que en varios ítems los hombres son mejor evaluados los últimos años, sin embargo en términos generales las mujeres siguen siendo mejor evaluadas en ambos niveles, pero la brecha disminuye tanto que su diferencia es casi inexistente o no significativa en los últimos años.

Respecto a la triple interacción que se da entre “Sexo del alumno”, “Sexo del profesor” y “Nivel de estudios”, podemos ver simultáneamente que las diferencias encontradas no son constantes ni por los años de estudio, ni por el sexo del profesor, así como tampoco por sexo del alumno, lo que nos permite entender mejor el sentido inverso que asumen las evaluaciones de los profesores hombres y profesoras mujeres.

En términos generales pudimos apreciar que los puntajes más bajos de los primeros años de estudio corresponden a la interacción “profesor hombre-estudiante hombre”, es decir, los profesores son evaluados mal sobre todo por los alumnos hombres los primeros años, mientras que los últimos años los puntajes más bajos se definieron por la interacción “profesor mujer-alumno hombre”, vale decir, los alumnos hombres invierten su lógica de evaluación a lo largo del ciclo de estudios.

Por otro lado, los primeros años los mejor evaluados estuvieron definidos por la interacción “profesoras mujeres-alumnas mujeres”, mientras que en los últimos años los mejor evaluados estuvieron definidos por la interacción “profesor hombre- alumnas mujeres”.

En ese sentido se mantiene la lógica respecto a que los hombres evalúan peor y las mujeres mejor, lo que irá variando será el género de los profesores mejor y peor evaluados según los distintos años de estudio.

En el caso de las alumnas mujeres, se podría decir que hay un traspaso de “lealtad” o “popularidad”, donde en los últimos años dejan de evaluar mejor a las mujeres y empiezan a evaluar mejor a los hombres. En algunos casos esta tendencia es más leve (“Entrega pautas corrección”), en otros casos es más fuerte (“Acerca conocimiento mediante ejemplos”). Lo cierto es que en términos netos, las profesoras mujeres pierden la ventaja que tenían, e incluso la relación entre profesor hombre y profesor mujer se invierte en la mayoría de los ítems. En otras palabras, los hombres serán en general mejor evaluados los últimos años, pero quienes evalúan mejor a sus profesores hombres en los últimos años serán sobre todo las alumnas mujeres. Esto puede explicar en parte el aumento general en los puntajes asignados a profesores los últimos años que hablamos previamente, es decir, el aumento en los puntajes asignados a profesores hombres, principalmente por estudiantes mujeres, fue lo suficientemente importante como para arrastrar a todos los puntajes en general a ser en promedio mayores los últimos años respecto a los primeros.

Una excepción importante a esta tendencia serán los ítems relacionados a conocimiento, específicamente “Dominio de contenidos” y “Presentar variados enfoques”. En ambos ítems ya habíamos corroborado que los hombres son siempre mejor evaluados. Acá nuevamente observamos que los profesores hombres son mejor evaluados los primeros años y los últimos, pero vemos que en ambos casos quienes mejor los evalúan son los estudiantes de sexo femenino. Por otro lado, las profesoras mujeres que son peor evaluadas los primeros y los últimos años, en ambos casos quienes peor las evalúan son estudiantes del sexo masculino. Sin embargo la diferencia no es tan pronunciada entre el sexo de los profesores los primeros años, y se hace más amplia los últimos, sobre todo porque el puntaje asignado por parte de los estudiantes hombres hacia las profesoras mujeres disminuye bruscamente los últimos años de estudio.

Por lo tanto se acepta la tercera hipótesis respecto a que los puntajes obtenidos por los docentes en la evaluación que realizan los estudiantes en la Universidad de Chile, varían

significativamente en relación a la interacción entre el género del profesor, el género del alumno y el nivel o año cursado por el alumno.

En suma, la constatación de que las profesoras mujeres tienden a ser mejor evaluadas en ítems asociados a la responsabilidad mientras que los profesores hombres tienden a ser mejor evaluados en ítems relativos al desempeño, sugiere que efectivamente hay diferencias de género, pero la pregunta respecto a la procedencia de esas diferencias muy difícilmente se podrá responder desde un cuestionario cuantitativo. De acuerdo a la literatura de género revisada, muy posiblemente las diferencias encontradas se deben a un contexto institucional, donde las condiciones para realizar satisfactoriamente el trabajo académico no serían las mismas para las mujeres que tienen que cumplir con patrones diferentes de educación, capacitación y desempeño para competir con sus colegas varones, y así neutralizar el coeficiente simbólico negativo con el que son asociadas. (Berríos, 2005:355). Por otro lado puede ser que las predisposiciones vengan por parte de los estudiantes, donde las instituciones de educación son en sí mismas mecanismos de reproducción cultural, social y valórica por lo que se repiten valores de género existentes en la sociedad (Bourdieu y Passeron, 1995).

Por lo tanto, la pregunta de si estas diferencias serán reales o serán atribuciones no se podrá responder desde un estudio cuantitativo, sino por medio de estudios longitudinales mixtos, que estudien también de manera cualitativa el proceso a lo largo del tiempo. De manera tal que, por lo menos desde un ámbito cuantitativo, no podremos establecer de manera tajante si las mujeres realmente tienen mayor empatía y responsabilidad, o son atribuciones que se le hacen, pero si podemos establecer que ambos sexos tienen patrones de evaluación diferenciados y estadísticamente significativos.

Estos hallazgos nos llevan de vuelta al tema de la utilidad y validez de los cuestionarios de evaluación docente. Tenemos por un lado importantes antecedentes de estudios que han llegado al consenso de que los factores extra-clase juegan un papel menor en las clasificaciones del estudiante sobre su propia instrucción (Marsh, 1987; Murray, 1984). Sin embargo también revisamos estudios de la última década que han ido dejando al descubierto lo significativo que resultan los factores extra clase o predisposiciones en las

puntuaciones finales que los estudiantes hacen de sus profesores (Onwuegbuzie, 2007, Basow, 1995, 2000).

Nuestra impresión, a partir de los resultados, es que si bien estos hallazgos pueden llegar a sugerir debilidades en la validez de estos cuestionarios, sobre todo en lo relativo a la validez del constructo (Onwuegbuzie, 2007), los tamaños de los efectos encontrados no son lo suficientemente grandes como para invalidar completamente el uso de ellos. Sin embargo, el que los tamaños de los efectos no sean lo suficientemente grandes como para invalidar el constructo, no quiere decir que no encontremos importantes estas diferencias, sobre todo porque este tipo de estudio permitiría identificar de manera más específica el ámbito donde estos sesgos estarían operando. Siguiendo las observaciones hechas por Basow (1995), vimos que los ítems presentados en los puntajes agregados no muestran patrones muy específicos respecto a las diferencias de género, más allá de algunos pocos ítems. Sin embargo al ver estos ítems en interacción con otras variables aparecieron una serie de patrones de género donde, si bien los tamaños de los efectos fueron pequeños como es común en ámbitos de la educación, si resultaron ser significativos por el tamaño de la muestra y por la consistencia que estos mostraron como tendencia general.

Lo anterior nos sugiere que develar el impacto de las variables de género desde estudios cuantitativos requiere estudiar las interacciones, ya que estas variables suelen operar en el micro nivel, como descriptores específicos. De esta manera se abren preguntas interesantes para futuras investigaciones, especialmente respecto a cómo estaría operando la variable ciclo de estudios que es la que justamente nos permitió sacar las conclusiones más interesantes. Una posibilidad hipotética es que la variable de ciclo de estudio tenga que ver con introyectar patrones disciplinarios, es decir, entender el paso de los años como un proceso de socialización disciplinaria. Esto explicaría por qué las mujeres que tendían a evaluar más desde la empatía los primeros años, aprenden a evaluar con criterios más disciplinarios, de una disciplina más bien masculina (en otras palabras, aprenden a evaluar como los hombres) como producto de un proceso de socialización disciplinaria.

Nuestra consideración a partir de estos resultados, y frente al avance que suponen los cuestionarios de evaluación docente como un instrumento para tomar decisiones políticas, es que se hace necesario interpretar sus resultados tomando en cuenta las variables

contextuales que inciden en mayor o menor medida en los puntajes finales de los profesores. En otras palabras, se hace necesario realizar y difundir estudios que den cuenta de cómo los patrones de género u otros factores estarían operando, para de esa manera ser más cautos a la hora de interpretar y ponderar los resultados de la evaluación.

Es importante mencionar que si bien los resultados son consistentes y válidos dentro de la facultad de ciencias sociales, estos no son necesariamente válidos en otras facultades o áreas académicas. Esto es así ya que, como vimos en la revisión bibliográfica, diversos estudios han demostrado que las diferencias disciplinarias son de suma importancia a la hora de hablar de evaluación docente (Hills, Naegle y Bartkus, 1997), lo que lleva a que los resultados de este estudio se apliquen como una descripción específica de lo que ocurre en cierto tipo de campos académicos, vale decir, en el campo de las ciencias sociales. Por lo mismo sería interesante replicar este tipo de estudio en otras facultades, rescatando la relevancia que asumen los patrones de género en la evaluación docente y la importancia de buscar estos patrones en las interacciones específicas con otras variables contextuales que pueden estar modulando los resultados. Esto permitiría, además de generar un valioso conocimiento para una mejor aplicación de los cuestionarios dentro de las disciplinas específicas, realizar interesantes comparaciones entre distintas facultades y áreas académicas. Esto último permitiría reflexionar y generar un mejor conocimiento acerca de la naturaleza diferente de las disciplinas, y más importante aún, permitiría observar y describir cómo ciertas estructuras propiamente culturales y sociales, como el imaginario de lo femenino y lo masculino, se estaría reproduciendo de manera tácita en las distintas áreas académicas.

En relación a lo anterior, todos los resultados que hemos descrito en esta investigación nos llevan finalmente a reflexionar sobre estos aspectos muchas veces oscurecidos por criterios más generales, y es que en el mundo de la universidad y de la academia las diferencias entre lo femenino y lo masculino estarían operando, por lo menos al nivel de los estudiantes, lo que impone nuevos desafíos respecto a cómo generar dentro de la misma universidad los procesos de autorreflexión necesarios para repensar las diferencias de género y poder así describir y explicitar cómo estos patrones de distinción estarían efectivamente actuando e incidiendo en los distintos espacios académicos.

Bibliografía

Abrami P. C; Sylvia d'Apollonia (1997). Navigating Student Ratings of Instruction. *American Psychological Association*, 11, 1198-1208,

Aleamoni, L. M. (1999). Student Rating Myths Versus Research Facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13:2 153.

Aleamoni & Hexner, (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*, 9, 67-84.

Barbero, M.I. (2003). *Psicometría II, métodos de elaboración de escalas*. Madrid: UNED.

Basow, S. A. (1995). Students evaluations of college professors: When gender matters. *Journal of educational psychology*, 87

Basow, S. A. (1998). Student evaluations: Gender bias and teaching styles. In L. H. Collins, Chrisler, J.C., & Quina, K. (Eds.), *Career strategies for women in academe: Arming Athena*. 135-156. Thousand Oaks, CA:

Basow, S. A. (2000). Best and Worst Professors: Gender Patterns in Students' Choices, *Sex Roles*, 43

Basow, Phelan & Capotosto (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly*, 30, 25–35.

Bourdieu Pierre (2003) *El oficio de científico: ciencia de la ciencia y reflexividad*. Anagrama. España

Bourdieu, P., & Passeron, J. (1977). *Reproduction in education society and culture* Thousand Oaks, CA: Sage.

Berríos, P. (2005). El sistema de prestigio en las universidades y el rol que ocupan las mujeres en el mundo académico, *Revista Calidad en la Educación*, 23, 351-361.

Becher, T (2001). Tribus y territorios académicos. La indagación intelectual y las culturas de las disciplinas. Gedisa, Barcelona

Cohen, P. A. (1990). Validity of Student Ratings of Instruction: What We Know and What We Do Not, *Journal of Educational Psychology*, 2,219-231

Cohen, J, (1988). Statistical Power Analysis for the Behavioral Sciences. Erlbaum

Feldman, K. A. (1983). “Seniority and experience of college teachers as related to evaluations they receive from students”, en *Research in Higher Education*, 18, 3-123.

Feldman, K. A. (1984). Class size and college students’ evaluations of teachers and courses: A closer look, en *Research in Higher Education*, 21, 45-116.

Feldman K.A (1986) The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristic a review and synthesis, *Reseach in Higher education*, 24

Feldman, K. A. (1987). Research productivity and scholarly accomplishments of college teachers as related to their instructional effectiveness, *Research in Higher Education*, 26, 227-298.

Feldman. K. A. (1978). Course characteristics and college students’ ratings of their teachers and courses: What we know and what we don’t, en *Research in Higher Education*, 9, 199-242.

Feldman, K. A. (1993). College students’ views of male and female teachers: Part II-evidence from students’ evaluations of their classroom teachers. *Research in Higher Education*, 34, 151-211.

Fuhrmann B. G; Geyer A. (2003). Students’ Evaluation of Teachers and Instructional Quality—Analysis of Relevant Factors Based on Empirical Evaluation Research. *Assessment & Evaluation in Higher Education*, Vol. 28, No. 3.

Garduño J. M. G. (2000). ¿Qué factores extraclase o sesgos que afectan la evaluación docente en la educación superior? *Revista Mexicana de investigación educativa, volumen 5, número 10*.

Garduño J. M. G. (2008). El proceso perverso de la evaluación de la docencia en las universidades: Un balance inicial y apuntes para mejorarlo. *Reencuentro 53, 9-19*

Hills S. B; Naegle N; Bartkus K. R. (1997). How Important Are Items on a Student Evaluation? A Study of Item Saliency. *Journal of Education for Business, 5, 297-303*

Jiménez, Herrera. & Rojas (2002). Fuentes de varianza e índices de varianza explicada en las ciencias del movimiento humano. *Revista de Ciencias del Ejercicio y la Salud, 2, 70-74*.

Lamadrid, Silvia (2012) *Desarrollo académico y políticas de género en las universidades*. Ponencia presentada e XIV Encuentro de Género, Guadalajara, Mexico.

Marsh H. W; Roche L. A (1997). Making Students' Evaluations of Teaching Effectiveness Effective. *American Psychologist, 11, 1187-1197*.

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness from different academic settings and their relation to student/course/instructor characteristics. , *Journal of Educational Psychology, 75, 150-166*.

Marsh H. W. (1984). Students evaluation of university teaching: Dimensionality, reliability, validity, potencial biases, and utility. *Journal of educational Psychology, 76,*

Marsh, H. W.; Overall, J. U. y Kesler, S. P. (1979). Class size, students' evaluations, and instructional effectiveness.. *American Educational Research Journal, 16, 57-69*.

Morales Vallejo P (2011). El tamaño del efecto (effect size): análisis complementarios al contraste de medias Estadística aplicada a las Ciencias Sociales. Universidad Pontificia Comillas, Madrid, Facultad de Ciencias Humanas y Sociales.

<<http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oDeLEfecto.pdf>>

(Consultado el 29 de marzo 2012).

Neumann, Y. y Neumann, L. (1983). Characteristics of academic areas and students' evaluation of instruction, *Research in Higher Education* 19, 323-334.

Onwuegbuzie, A. J., Witcher, A. E., Collins, K. M. T., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed methods analysis. *American Educational Research Journal*, 44, 113-160.

Olavarría J (2011). La participación y el liderazgo de las mujeres que participan en la educación superior (IES) en las Américas. Informe final Red Interamericana de formación en mujeres, género y desarrollo con equidad (RIF GED). <http://www.ouiiohe.org/webcolam/rifged/iu/search.php?q=olavarr%C3%ADa> (Recuperado el 2 de Diciembre del 2012)

Reichardt, C.S. & Cook, T.D (1986). Hacia una superación del enfrentamiento entre los métodos cualitativos y los cuantitativos. Madrid: Morata.

Smith; Yoo; Farr; Salmon; & Miller (2007) The Influence of Student Sex and Instructor Sex on Student Ratings of Instructors: Results from a college of communication. *Women's Studies in Communication*, 30

Rizo V; Coello G. (2006). Percepciones estudiantiles y su influencia en la evaluación del profesorado. Un caso en la Universidad Autónoma de Occidente, Cali-Colombia. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_3.htm (Consultado el 6 de Nov. 2012)

Spencer K. J; Pedhazur L. (2002). Student Evaluations and gendered expectations: What we can't count can hurt us. *Assessment & Evaluation in Higher Education*, Vol. 27, No. 5.

Sprague J; Massoni K. (2005). Student Evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, Vol. 53.

Smith R. A; Cranton P. A. (1992) Students' perceptions of teaching skills and overall effectiveness across instructional settings. *Research in Higher education*, 6, 747-764

Saracostti, M. (2006) Mujeres en la alta dirección de educación superior: posibilidades, tensiones y nuevas interrogantes. *Revista Calidad en la educación*, 25, 243-259.

Tatro, C.N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education*, 3, 169-173.

Vásquez Rizo, F. E. y Gabalán Coello, J. (2006). Percepciones estudiantiles y su influencia en la evaluación del profesorado. Un caso en la Universidad Autónoma de Occidente, Cali-Colombia. Revista *RELIEVE*, 2, 219-245

Valentine J. & COOPER, H. (2003). Effect Size Substantive Interpretation Guidelines: Issues in the Interpretation of Effect Sizes.

<<http://ies.ed.gov/ncee/wwc/document.aspx?sid=1&pid=2>> (consultado 6, Dic. 2011).

Vucht Tijssen, L. van, & E. de Weert (2005) From Erudition to Academic Competence. in: *Revista española de pedagogía*, 230, 123 – 145.

Softwares estadísticos utilizados

LISREL 8.80 for Windows. Jöreskog & Sörbom (2006). Disponible en: <http://www.ssicentral.com/> [descargado el 4, mayo 2012]

Ncalculators. Disponible en <http://es.ncalculators.com/math-calculators.html> [utilizado el 9, junio 2012]

IBM SPSS Statistics, versión 19 (IBM, EE.UU.) para Windows. Disponible en: <http://www.spss.com>. [Accedido el: 20 mayo 2011]

FACTOR 7.2 (Lorenzo-Seva & Ferrando, 2006). Disponible en <http://psico.fcep.urv.es/utilitats/factor/>[descargado el: 20 mayo 2011]

Anexo I: Pauta de evaluación docente

I.- Características de la Asignatura

Nombre de la Asignatura: _____

	0% al 25%	26% al 50%	51% al 75%	75% al 100%		
1.- ¿En cuál de los siguientes rangos estuvo tu porcentaje de asistencia al curso?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
2.- ¿En qué rango se encontró tu promedio final del curso?	1 a 3,9 <input type="checkbox"/>	4,0 a 4,9 <input type="checkbox"/>	5,0 a 5,9 <input type="checkbox"/>	6,0 a 6,9 <input type="checkbox"/>		
3.- ¿En qué oportunidad rendiste el curso?	Primera <input type="checkbox"/>	Segunda <input type="checkbox"/>	Tercera o más <input type="checkbox"/>			
4.- ¿Cómo crees que fue el tiempo de dedicación que te requirió el curso?	Demasiado <input type="checkbox"/>	Adecuado <input type="checkbox"/>	Muy poco <input type="checkbox"/>			
5.- ¿Cómo crees que fue el nivel de exigencia del curso?	Demasiado alto <input type="checkbox"/>	Alto, pero adecuado <input type="checkbox"/>	Bajo <input type="checkbox"/>	Demasiado Bajo <input type="checkbox"/>		
6.- ¿Existe superposición entre los contenidos de este curso y otros anteriores?	Si, hay demasiada superposición <input type="checkbox"/>	Si, pero la superposición es útil <input type="checkbox"/>	No hay superposición <input type="checkbox"/>			
7.- ¿Cómo fue la calidad del curso?	Muy alta <input type="checkbox"/>	Alta <input type="checkbox"/>	Regular <input type="checkbox"/>	Baja <input type="checkbox"/>	Muy Baja <input type="checkbox"/>	
8.- ¿Cómo fue la relación entre el programa y las actividades desarrolladas?	Muy alta <input type="checkbox"/>	Alta <input type="checkbox"/>	Regular <input type="checkbox"/>	Baja <input type="checkbox"/>	Muy Baja <input type="checkbox"/>	
9.- ¿Cuál es la utilidad que crees tiene este curso respecto a tu formación?	Muy alta <input type="checkbox"/>	Alta <input type="checkbox"/>	Regular <input type="checkbox"/>	Baja <input type="checkbox"/>	Muy Baja <input type="checkbox"/>	
10.- ¿Este curso fue impartido por un profesor o un equipo?	Fue impartido por un profesor <input type="checkbox"/>		Fue impartido por un equipo <input type="checkbox"/>			
11.- Si es que el curso fue	Muy buena <input type="checkbox"/>	Buena <input type="checkbox"/>	Regular <input type="checkbox"/>	Mala <input type="checkbox"/>	Muy Mala <input type="checkbox"/>	No aplicable <input type="checkbox"/>

impartido por un equipo, ¿Cómo fue la coordinación entre los profesores?

II.- Evaluación de cada profesor:

Nombre del Profesor: _____

Sobre el profesor, ¿Cuán de acuerdo estás con las siguientes afirmaciones?:

	Muy de acuerdo	De acuerdo	En desacuerdo	Muy en desacuerdo	No aplicable
1.- El profesor inició y terminó puntualmente su clase.					
2.- El profesor fue puntual en la entrega de evaluaciones.					
3.- El profesor asistió regularmente a clases.					
4.- El profesor domina los contenidos de su clase.					
5.- El profesor se comunica con claridad (al responder preguntas, explicar contenidos, etc.).					
6.- Los contenidos expuestos por el profesor están actualizados y tienen relación con la realidad actual.					
7.- El profesor presentó variados enfoques teóricos referentes a la materia.					
8.- El profesor utiliza estrategias de enseñanza que estimulan la participación de los estudiantes.					
9.- El profesor demuestra compromiso con el desarrollo del aprendizaje de los estudiantes.					
10.- El profesor muestra disposición a enseñar y aclarar dudas, tanto dentro como fuera de la hora de clases.					
11.- El profesor acerca al estudiante al conocimiento de la disciplina a través de situaciones concretas o ejemplos.					
12.- El profesor logra un trato equitativo y respetuoso en su relación con los estudiantes.					
13.- El profesor es consistente entre lo enseñado y lo exigido en las evaluaciones.					
14.- El profesor prepara materiales útiles y pertinentes para el aprendizaje.					
15.- El profesor entrega el programa oportunamente (al inicio del semestre).					

16.- El profesor entregó pautas claras de corrección de pruebas y trabajos.					
---	--	--	--	--	--

17.- En síntesis y en términos generales, ¿Cómo calificarías al docente?:

Muy bueno Bueno Regular Malo Muy malo

18.- ¿Tienes algún comentario general que hacer al curso o al profesor?.

Anexo II. Medidas de distribución.

Medidas de simetría

Nombre de variable	N	Mínimo	Máximo	Media	Desv. típ.	Asimetría		Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error típico	Estadístico	Error típico
Puntualidad inicio - término clases	8572	1	4	3.13	.820	-.838	.026	.342	.053
Puntualidad entrega evaluaciones	8367	1	4	2.90	.898	-.542	.027	-.411	.054
Asistencia regular a clases	8527	1	4	3.30	.757	-1.048	.027	1.021	.053
Dominio contenidos	8573	1	4	3.44	.688	-1.192	.026	1.498	.053
Comunicación clara	8566	1	4	3.14	.836	-.783	.026	.063	.053
Actualización y relación con realidad actual	8387	1	4	3.26	.726	-.918	.027	.983	.053
Presentar variados enfoques	7962	1	4	3.00	.848	-.598	.027	-.189	.055
Estimula participación	8490	1	4	2.81	.926	-.383	.027	-.693	.053
Compromiso con aprendizaje	8522	1	4	3.05	.835	-.750	.027	.166	.053
Disposición a enseñar	8469	1	4	3.19	.774	-.904	.027	.736	.053
Acerca conocimiento mediante ejemplos	8470	1	4	3.20	.735	-.804	.027	.687	.053
Trato equitativo	8553	1	4	3.22	.774	-.979	.026	.934	.053
Consistencia entre enseñado y exigido	8410	1	4	3.06	.832	-.782	.027	.262	.053
Materiales útiles	8347	1	4	2.95	.846	-.555	.027	-.216	.054
Entrega oportuna programa	8242	1	4	3.22	.773	-.955	.027	.821	.054
Entrega pautas corrección	8264	1	4	2.77	.936	-.359	.027	-.735	.054