

Adaptación y Validación Preliminar del SPQ y el CEQ Para el Estudio de la Formación en Docencia Universitaria en el Contexto Chileno

Preliminary Adaptation and Validation of the SPQ and the CEQ for the Study of Teaching Development Programmes in a Chilean University Context

Jorge Marchant, Jaime Fauré y Nicole Abricot
Universidad de Santiago de Chile

El estudio tuvo como objetivo adaptar y aportar evidencias preliminares de validez del *Course Experience Questionnaire* (CEQ) y el *Study Process Questionnaire* (SPQ) en el contexto de la formación en docencia universitaria. Se analizaron evidencias de validez de contenido mediante el juicio de 6 expertos, estructura interna a través de análisis factorial confirmatorio, correlación entre los instrumentos y consistencia interna de ambos instrumentos, a través de los coeficientes alfa y omega. La muestra por conveniencia estuvo compuesta por 686 estudiantes de una universidad de la Región Metropolitana de Chile. Se obtuvieron versiones revisadas y adecuadas lingüísticamente de los instrumentos, alto grado de acuerdo entre jueces sobre su contenido y evidencias a favor de la estructura de 2 factores del SPQ y 4 del CEQ. También se encontraron correlaciones teóricamente consistentes según el modelo 3P, donde, por ejemplo, aquellos estudiantes con un enfoque profundo de aprendizaje percibieron una buena enseñanza y metas y objetivos claros. Se discute la posibilidad de desarrollar argumentos a favor de la utilización del CEQ y el SPQ a partir de otro tipo de evidencias.

Palabras clave: validación, consistencia interna, SPQ, CEQ, Chile

The aim of this study was to adapt the *Course Experience Questionnaire* (CEQ) and the *Study Process Questionnaire* (SPQ) and to provide preliminary evidence of their validity in the context of university teaching developmental programs. Content validity evidence was analyzed by 6 expert raters, internal structure was checked via confirmatory factor analysis, and internal consistency and correlation between the instruments were examined via alpha and omega coefficients. The convenience sample was composed of 686 students from a university located in the Metropolitan Region of Chile. Revised and linguistically adapted versions of the instruments were obtained. A high degree of agreement among the raters was observed regarding the contents of the instruments. The evidence obtained supported the 2-factor structure of the SPQ and the 4-factor structure of the CEQ. Also, theoretically consistent correlations were found according to the 3P model. For example, students with a deep learning approach perceived good teaching and clear objectives and goals. The possibility to develop arguments for the use of the CEQ and the SPQ based on other types of evidence is discussed.

Keywords: validation, reliability, SPQ, CEQ, Chile

Durante las últimas décadas, la educación superior ha pasado por una serie de cambios que han tenido como eje el énfasis en la mejora de los procesos de enseñanza y aprendizaje y el aseguramiento de la calidad (Longden, 2010). Por una parte, se ha pasado de un modelo educativo centrado principalmente en el docente a un modelo en el que se valoriza el papel protagónico y activo del alumno en la construcción de aprendizajes. Por otra, se ha puesto el énfasis en el aseguramiento de la calidad, lo que ha desembocado en la necesidad de contar con modelos conceptuales que permitan poner a prueba la calidad de las intervenciones realizadas sobre los fenómenos educativos, mediante instrumentos que faciliten la extracción de conclusiones válidas y confiables (Ramsden, 2003).

Jorge Cristián Marchant Mayol, Jaime Ignacio Fauré Niños y Nicole Edith Abricot Marchant, Unidad de Innovación Educativa, Universidad de Santiago de Chile.

Esta investigación ha sido financiada gracias al aporte del Ministerio de Educación de Chile y se encuentra bajo el alero de un Proyecto FIAC2, código USA1307.

La correspondencia relativa a este artículo debe ser dirigida a Jorge Marchant, Unidad de Innovación Educativa, Universidad de Santiago de Chile, Las Sophoras 175, Estación Central, Santiago, Chile. E-mail: jorge.marchant@usach.cl

En este contexto, la docencia adquiere una prioridad estratégica para las instituciones de educación superior, especialmente si se considera que diferentes estudios han sugerido que corresponde a una variable altamente explicativa de la calidad institucional (Cid-Sabucedo, Pérez-Abellás & Zabalza, 2009) y, más importante aún, constituye un factor que incide en el tipo de aprendizajes que alcanzan los estudiantes (Trigwell, Prosser & Waterhouse, 1999).

Esta marcada preocupación por comprender la calidad de la docencia universitaria ha desembocado en la emergencia de diferentes modelos explicativos. Coll y Solé (2001) y Shulman (1986/1989) coinciden en que los estudios que han prevalecido durante el siglo pasado se han centrado en analizar el comportamiento observable del profesor, bajo el supuesto de que este comportamiento es la variable más importante si se desea explicar el rendimiento académico de los estudiantes. Desde esta perspectiva, se han desarrollado los llamados *modelos de enseñanza eficaz* (Zabalza, 2003).

Tanto Coll y Solé (2001) como Shulman (1986/1989) han criticado los modelos centrados en el comportamiento del profesor, planteando que implican una excesiva simplificación y generalización de la enseñanza. Shulman (1986/1989) llegó a denominar esta perspectiva como *paradigma proceso-producto*, ya que en esta lógica otras variables de la enseñanza y el aprendizaje como, por ejemplo el contexto de aula o los contenidos de la enseñanza, resultan prácticamente invisibilizadas. Más aun, al considerar este modelo en el análisis de una enseñanza de calidad (Coll & Sánchez, 2008; Coll & Solé, 2001), es posible señalar que no hay una mejor forma de enseñar, ya que una buena docencia puede conseguirse a través de distintos métodos que consideren las necesidades de aprendizaje particulares de los estudiantes (Ramsden, 2003).

Este último desafío es el que intenta abordar el *modelo 3P*, planteado inicialmente por Biggs en 1987, a partir del modelo desarrollado por Dunkin y Biddle en 1974, y adaptado posteriormente por Prosser y Trigwell en 2006. En términos explicativos, este modelo señala tres puntos en que se sitúan los factores que interactúan en los procesos de enseñanza y aprendizaje: pronóstico, antes de que se produzca el aprendizaje; proceso, durante el aprendizaje, y producto (Biggs, 1987; Prosser & Trigwell, 2006).

Prosser y Trigwell (2006) plantean que los factores asociados al pronóstico son de tres tipos: dependientes de los estudiantes, dependientes del contexto de enseñanza y dependientes del docente. En este sentido, los estudiantes y los profesores se incorporan a cualquier situación de aprendizaje con experiencias previas (incluidas sus experiencias de aprendizaje y enseñanza, origen socioeconómico, concepciones sobre qué significa aprender y enseñar, entre otras) que forman su marco de referencia contra el cual interpretan la situación en la que se encuentran (Pichardo Martínez, García Berbén, De la Fuente Arias & Justicia Justicia, 2007).

Estos factores interactúan en el nivel del proceso, condicionando las percepciones y enfoques asumidos por estudiantes y docentes. En una situación de aprendizaje concreta, los estudiantes pueden adoptar formas profundas o superficiales para aprender. De acuerdo a Biggs (1999), cuando adoptan un enfoque profundo, los estudiantes se enfocan en el significado de los contenidos, relacionándolos con sus conocimientos previos e integrando aspectos teóricos y prácticos novedosos. Por su parte, cuando adoptan un enfoque superficial, los estudiantes realizan las tareas con el mínimo esfuerzo, focalizándose en el desarrollo de actividades de bajo nivel cognitivo (Biggs, 1999). Estos enfoques describen la manera en que cada estudiante se relaciona con su ambiente de enseñanza y aprendizaje, por lo que no deben pensarse como características fijas ni estables en ellos, sino como estrategias adoptadas en función del contexto de aprendizaje percibido (Biggs, 1999; Ramsden, 2003).

Por otro lado, los enfoques de aprendizaje de los estudiantes se relacionan con sus experiencias de aprendizaje. Una característica interesante del modelo 3P es que plantea que los aprendizajes que alcanzan los estudiantes responden a cómo perciben algunos elementos del contexto concreto en el que se encuentran, como la buena docencia, las metas y objetivos claros, la evaluación apropiada o la carga de trabajo apropiada, entre otros (Ginns & Ellis, 2009; Webster, Chan, Prosser & Watkins, 2009). Más específicamente, la investigación revela que la percepción de buena docencia y metas y objetivos claros presentan correlaciones directas con el uso de un enfoque profundo de aprendizaje, a diferencia de la percepción sobre la carga académica de trabajo y la evaluación, que presentan correlaciones con el aprendizaje superficial (Ramsden, 2003). En otras palabras, dichas experiencias de aprendizaje pueden condicionar en una u otra dirección el enfoque u orientación que asumen los estudiantes. Asimismo, la evidencia propone que la buena docencia no se relaciona con el aprendizaje superficial, mientras que la evaluación apropiada apenas presenta correlaciones con el aprendizaje profundo (González, Montenegro, López, Munita & Collao, 2011).

Finalmente, a nivel del producto, tanto los resultados de la enseñanza como del aprendizaje estarán determinados por la interacción entre todos los factores mencionados anteriormente. Es el caso, por ejemplo, del rendimiento académico (Doménech Betoret, Jara Jiménez & Rosel Remírez, 2004).

Todo lo anterior queda resumido en la Figura 1.

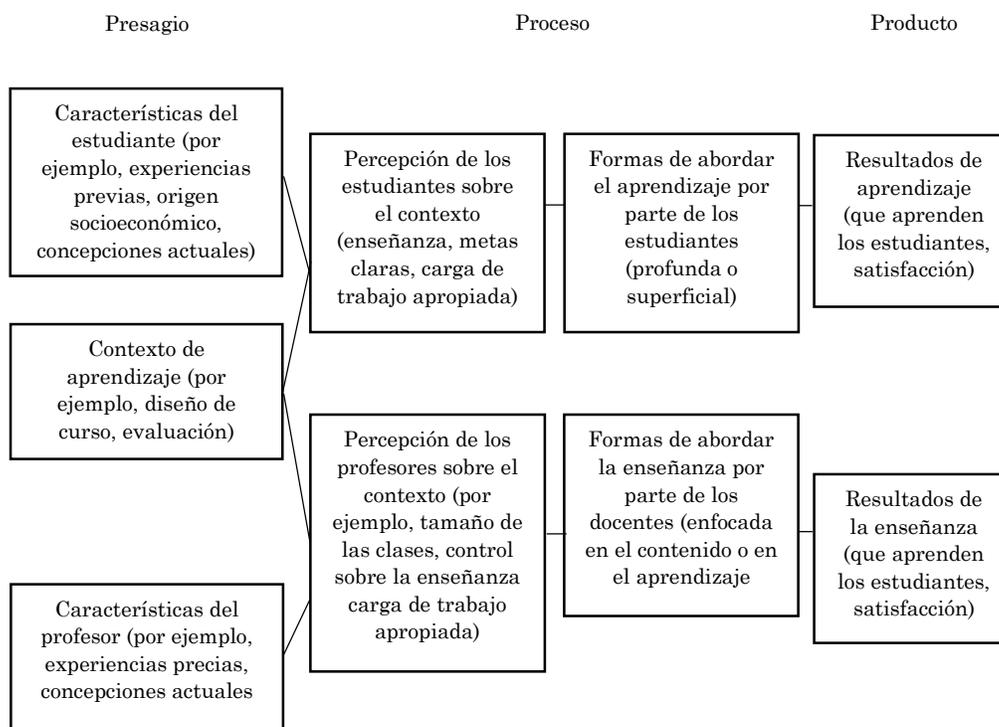


Figura 1. Modelo 3P. Extraído de “Calidad del Aprendizaje Universitario” de Biggs (2008).

Si bien parece evidente, es necesario explicitar que el uso del modelo 3P supone darle importancia no solo a los resultados de enseñanza y aprendizaje, sino también a lo que ocurre antes y durante los procesos educativos (Biggs, 1999). Esto coincide con los cambios de énfasis en la educación superior mencionados previamente, que han orientado el foco institucional hacia la consecución de aprendizajes de calidad por parte de los estudiantes (Longden, 2010). En esta línea, las experiencias y enfoques de aprendizaje de los estudiantes adquieren una relevancia especial, puesto que se relacionan de manera directa con la calidad de los aprendizajes (Biggs, 1987; Marton, Hounsell & Entwistle, 1984).

Lo anterior ha producido que diferentes equipos de trabajo hayan creado y desarrollado instrumentos de medición cuantitativos, que permiten evaluar estas variables propuestas por este modelo con diferentes propósitos. El modelo 3P ha tenido una dilatada difusión y ha sido ampliamente estudiado en diferentes contextos, por lo que actualmente constituye un referente teórico y metodológico en diferentes países y organizaciones en el mundo, tanto para investigar los procesos de enseñanza y aprendizaje en general como para evaluar asignaturas, programas universitarios completos y programas específicos de formación (De la Fuente Arias, Martínez Vicente, Peralta Sánchez & Berben, 2010; Munshi, Al-Rukban & Al-Hoqail, 2012).

Especialmente importante en este sentido es la línea de evaluación de los programas de formación de profesores (Fryer, Ginns, Walker & Nakao, 2012), en la que algunos estudios se han enfocado en el impacto de los programas únicamente sobre la enseñanza y el comportamiento docente (Postareff & Lindblom-Ylänne, 2008; Postareff, Lindblom-Ylänne & Nevgi, 2007; Stes, De Maeyer & Van Petegem, 2013), mientras otros han complementado esta información con la evaluación de la formación, a través de la percepción y los enfoques adoptados por los estudiantes (Gibbs & Coffey, 2004). A modo de resumen, uno de los hallazgos significativos a partir de esta línea de desarrollo consiste en que aquellos profesores que han recibido formación y logran

desarrollar estrategias centradas en los alumnos tienen más probabilidades de generar un abordaje profundo de aprendizaje en sus estudiantes, mientras que aquellos que centran sus estrategias en la transmisión de información tienen más probabilidades de tener estudiantes que usen abordajes superficiales.

El uso de herramientas cuantitativas en este contexto se justifica por la facilidad que ofrecen en la recolección de datos y en la presentación de resultados y su interpretación (Spooren, Brockx & Mortelmans, 2013). Si bien la lista de instrumentos es extensa, en la literatura especializada se reconoce el uso del *Study Process Questionnaire* (SPQ) y el *Course Experience Questionnaire* (CEQ). El SPQ fue construido por Biggs (1987) y es utilizado para evaluar la forma en que los estudiantes abordan su aprendizaje. En otras palabras, permite identificar si el estudiante ocupa en mayor medida un enfoque de aprendizaje profundo o un aprendizaje superficial (Prosser & Trigwell, 2001). Por su parte, el CEQ fue desarrollado por Ramsden (1991) con el objetivo de medir la percepción de los estudiantes de la calidad de la enseñanza en sus programas o cursos, a partir de diferentes escalas, entre las que se encuentran: calidad general de la docencia, carga de trabajo, metas y objetivos claros, adecuación de las evaluaciones y énfasis en la autonomía, entre otras.

Se han creado otros instrumentos para evaluar los enfoques de aprendizaje o las experiencias de aprendizaje, como el *Approaches and Study Skills Inventory for Students* (Entwistle, 1997) o el *Noel-Levitz Student Satisfaction Inventory* (Richardson, 2005), respectivamente. Sin embargo, estos instrumentos son utilizados con menor frecuencia dentro del marco de la formación docente por la comunidad académica y, por consiguiente, cuentan con menos evidencias de su pertinencia para la evaluación de los programas (González et al., 2011).

Si bien algunos estudios actuales han comenzado a indagar en los procesos de enseñanza y aprendizaje a través de la aplicación del modelo 3P con metodologías integradoras que incluyen una serie de estrategias multivariadas y multinivel para considerar cada elemento en su interrelación con los demás (Stes, De Maeyer, Gijbels & Van Petegem, 2012), los instrumentos que evalúan el eje de procesos del modelo 3P han sido construidos y desarrollados por separado. Asimismo, muchas veces son usados para responder a preguntas y necesidades específicas sobre las variables a las cuales se enfocan cada uno de ellos. Es por esto que en el presente artículo se prefirió mantener esta independencia en su uso e interpretación, y evaluar su validez para el estudio del impacto de la formación docente sobre los estudiantes.

Desde esta perspectiva, la validez se comprende como la condición que un test evidencie estar midiendo lo que debe estar midiendo, respecto de la finalidad para la cual fue construido. Por lo tanto, se trata de una evaluación de la coherencia e integridad de la interpretación y uso de los razonamientos y de la plausibilidad de sus inferencias y suposiciones (Hogan, 2004; Kane, 2013).

Al respecto, la literatura científica reconoce diversas estrategias para recoger evidencias que permiten elaborar un argumento sobre la validez de las puntuaciones. Según el *Standards for Educational and Psychological Testing*, las evidencias de validez pueden abordar el análisis del contenido, los procesos de respuesta, la estructura interna, la relación con otras variables y las consecuencias sobre los instrumentos (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 1999).

Por otra parte, la confiabilidad de los instrumentos se refiere a la precisión o estabilidad de las mediciones, y tiene relación con la posibilidad de generalizar los resultados obtenidos en una aplicación en la medida que son consistentes a través del tiempo y las circunstancias (Cronbach, 1951).

A continuación se describen las principales evidencias de validez del SPQ y el CEQ reportadas por la literatura internacional y chilena, específicamente con fines de investigación y evaluación de programas de formación docente.

Argumentos Internacionales

Study Process Questionnaire

En relación a las evidencias internacionales sobre la aplicación del SPQ, es posible mencionar que este instrumento ha sido adaptado a diferentes idiomas (e.g., árabe, japonés y español), usualmente con fines de investigación. En términos de las evidencias sobre su validez y confiabilidad, ha sido sometido a diferentes procesos de revisión de sus propiedades psicométricas, la mayor parte de las veces a través del análisis de la confiabilidad de sus puntuaciones y de la estructura interna de su composición factorial, lo que ha derivado en una versión denominada R-SPQ-2F (Stes et al., 2013). Los diferentes estudios, detallados en la Tabla 1,

señalan que el instrumento permite que las puntuaciones derivadas del mismo sean utilizadas considerando dos escalas: Aprendizaje Profundo y Superficial.

Tabla 1
Evidencias Psicométricas Internacionales del SPQ

Estudio	Análisis	Resultados
Munshi et al., (2012)	Análisis factorial exploratorio Consistencia interna	Modelo de dos factores: Aprendizaje Superficial y Aprendizaje Profundo $\alpha_{ap} = 0,90$; $\alpha_{as} = 0,93$
Fryer et al. (2012)	Análisis factorial confirmatorio Consistencia interna	Puesta a prueba de dos modelos, en la que el mejor ajuste estuvo en el modelo de cuatro factores: Motivo Profundo, Motivo Superficial, Estrategia Profunda y Estrategia Superficial $\alpha_{mp} = 0,56$; $\alpha_{ep} = 0,66$; $\alpha_{ms} = 0,45$; $\alpha_{es} = 0,36$; Escalas: $\alpha_{as} = 0,60$; $\alpha_{ap} = 0,76$
Stes et al. (2013)	Análisis factorial confirmatorio Análisis factorial exploratorio	No encuentran evidencia del ajuste del modelo teórico de dos factores con los datos recogidos Extracción por Máxima Verosimilitud con rotación Oblimín, obteniendo modelo de cuatro factores: Estudiar es Interesante, Gasto de Tiempo Extra, Mínimo Esfuerzo y Aprender Desde el Corazón
Socha & Sigler (2014)	Análisis factorial confirmatorio	Puesta a prueba de seis modelos, en la que el mejor ajuste estuvo en el modelo de dos factores: Aprendizaje Superficial y Aprendizaje Profundo

Course Experience Questionnaire

El CEQ fue construido en Australia por Ramsden (1991), con el objetivo de valorar la percepción de los estudiantes sobre la calidad de la enseñanza en sus programas completos de estudio. Sin embargo, posteriormente fue adaptado y validado para su utilización en otros países y en contextos de evaluación de un curso en particular (Richardson, 2005), y especialmente en el contexto de la formación docente (Fryer et al., 2012). En la Tabla 2 se resume la información disponible en la literatura científica, que apunta a la independencia de las escalas, las que, en consecuencia, han comenzado a trabajarse de forma independiente en los análisis.

Argumentos Nacionales

Para el caso chileno, es posible constatar que los estudios sobre esta temática son escasos y, en términos generales, los procesos de adaptación y validación a los que se han sometido requieren ser revisados.

Uno de los primeros estudios realizados en Chile fue conducido por González et al. (2011), quienes aplicaron el SPQ y el CEQ a una muestra de 1687 estudiantes tras un proceso de adaptación preliminar. En este caso, los autores no explicitan con claridad el procedimiento de adaptación, lo que no permite extraer conclusiones certeras sobre su pertinencia. Sin embargo, los análisis posteriores arrojaron estructuras teóricamente compatibles con el modelo 3P, aun cuando se reportaron algunas cargas cruzadas y cantidades bajas de varianza explicada en los análisis de Componentes Principales. Finalmente, para todas las escalas se obtuvieron estadísticos de consistencia interna apropiados.

Tabla 2
Evidencias Psicométricas Internacionales del CEQ

Estudio	Análisis	Resultados
Wilson, Lizzio & Ramsden (1997)	Análisis factorial exploratorio	Extracción por Componentes Principales con rotación Oblimín, obteniendo modelo de cinco factores que explican el 52% de la varianza
	Análisis factorial confirmatorio	Puesta a prueba de un modelo de cinco factores, en la que el mejor ajuste estuvo en el modelo de cinco factores: Buena Docencia, Metas y Objetivos Claros y Énfasis en la Independencia, Evaluación Apropiaada y Carga de Trabajo Apropiaada
	Consistencia Interna	$\alpha_{bd} = 0,80$; $\alpha_{moc} = 0,82$; $\alpha_{ea} = 0,73$; $\alpha_{ct} = 0,74$; $\alpha_{ei}^* = 0,68$
Lawless & Richardson (2002)	Análisis factorial exploratorio	Extracción por Factores Comunes, obteniendo modelo de siete factores que explican el 55,1% de la varianza
Byrne & Flood (2003)	Análisis factorial exploratorio	Extracción por Componentes Principales con rotación Oblimín, obteniendo modelo de cinco factores que explican el 53.5% de la varianza
	Consistencia Interna	$\alpha_{bd} = 0,76$; $\alpha_{moc} = 0,78$; $\alpha_{ea} = 0,73$; $\alpha_{ct} = 0,69$
	Correlación entre las escalas	$r_{bd-moc} = 0,38$; $r_{bd-ea} = 0,26$; $r_{bd-ct} = 0,22$; $r_{moc-ea} = 0,32$; $r_{moc-ct} = 0,08$; $r_{ea-ct} = 0,12$
Richardson (2005)	Análisis factorial exploratorio	Extracción por Componentes Principales, obteniendo modelo de cinco factores que explica el 56.5% de la varianza
	Consistencia Interna	$\alpha_{bd} = 0,79$; $\alpha_{moc} = 0,77$; $\alpha_{ea} = 0,47$; $\alpha_{ct} = 0,71$; $\alpha_{ei}^* = 0,55$
	Correlación entre las escalas	$r_{bd-moc} = 0,38$; $r_{bd-ea} = 0,44$; $r_{bd-ct} = 0,03$; $r_{moc-ea} = 0,22$; $r_{moc-ct} = 0,02$; $r_{ea-ct} = 0,12$; $r_{bd-ei} = 0,50$; $r_{moc-ei} = 0,40$; $r_{ct-ei} = 0,18$; $r_{ea-ei} = 0,40$
Ruiz & Schumacher (2008)**	Análisis factorial exploratorio	Extracción por Componentes Principales, obteniendo modelo de cinco factores que explican el 55.1% de la varianza
Fryer et al. (2012)	Análisis factorial exploratorio	Extracción por Ejes Principales con rotación Oblimín, obteniendo modelo de cuatro factores: Buena Docencia, Habilidades Generales, Evaluación Apropiaada y Carga de Trabajo Apropiaada
	Consistencia Interna	$\alpha_{bd} = 0,82$; $\alpha_{hg} = 0,73$; $\alpha_{ea} = 0,43$; $\alpha_{ct} = 0,61$
	Análisis factorial confirmatorio	Puesta a prueba de un modelo conjunto a SPQ, en la que el mejor ajuste estuvo en el modelo de cuatro factores: Buena Docencia, Habilidades Generales, Evaluación Apropiaada y Carga de Trabajo Apropiaada

* Énfasis en la Independencia

** Este estudio utilizó una versión del instrumento adaptada para el español en Colombia; sin embargo, esta versión no contó con un apropiado proceso de adaptación y cálculo de propiedades psicométricas.

Posteriormente, parte de este mismo equipo (González, Montenegro & López, 2012) adaptó cinco escalas del CEQ y revisó sus evidencias psicométricas, a través de la aplicación del instrumento a 325 estudiantes de ingeniería de una universidad pública chilena con fines investigativos. Las escalas adaptadas fueron Buena Docencia, Metas y Objetivos Claros, Evaluación Apropiaada y Carga de Trabajo Apropiaada, además de una escala de e-learning. Los autores tradujeron el instrumento siguiendo el método de *back translation* y realizaron un análisis de contenido mediante jueces expertos. Luego, buscaron evidencias de validez de constructo y de criterio. Respecto de las evidencias de validez de estructura interna, identificaron las escalas propuestas teóricamente a través de un análisis de Componentes Principales, en el que algunos ítems mostraron cargas cruzadas. Para evaluar la validez de criterio utilizaron como instrumento de comparación el SPQ, encontrando relaciones teóricamente plausibles. Sin embargo, el SPQ no estaba adaptado al contexto chileno, por lo que la evidencia pudiese no ser adecuada. Finalmente, obtuvieron consistencias internas, mediante el alfa de Cronbach, más bien bajas, que oscilaron entre 0,52 y 0,85. En resumen, es posible concluir que la investigación internacional y nacional aporta argumentos psicométricos para utilizar el CEQ y el SPQ en el análisis del eje de procesos del modelo 3P, especialmente en los contextos investigativos y de formación en docencia universitaria. Sin embargo, los estudios muestran en muchos casos evidencias contradictorias y/o debilidades metodológicas. En primer lugar, es necesario revisar y corregir los procesos de adaptación a nivel nacional para asegurar la calidad de los reactivos. En segundo lugar, se vuelve indispensable analizar con mayor detalle la estructura interna de los instrumentos. Por ejemplo, en el caso del SPQ algunos estudios reportan dos dimensiones (Munshi et al., 2012; Socha & Sigler, 2014), mientras otros proponen una estructura factorial de cuatro dimensiones (Fryer et al., 2012; Stes et al., 2013). En tercer lugar, los estudios

chilenos han reportado una consistencia interna baja para ambos instrumentos, por lo que también parece necesario indagar sobre dicho punto.

Entendiendo que existen diferentes fuentes y modelos metodológicos que permiten aportar evidencia sobre la validez de los instrumentos (AERA, APA & NCME, 1999) y a raíz del tipo de evidencias internacionales y nacionales reportadas anteriormente, el objetivo de la investigación fue aportar evidencias sobre la validez y consistencia interna de los cuestionarios SPQ y CEQ en el contexto chileno, a partir de un estudio aplicado al análisis de un curso de formación en docencia universitaria y su influencia en los procesos de enseñanza y aprendizaje, dentro del marco del modelo 3P.

Hipótesis

En la Tabla 3 se muestran las hipótesis que guiaron la búsqueda de evidencia empírica. En el caso de la evidencia de contenido, se esperaba que los jueces obtuvieran al menos 70% de acuerdo sobre la pertinencia de los contenidos de cada instrumento. Para el análisis de la estructura interna se esperaba rescatar las estructuras propuestas por la literatura, es decir, un modelo de dos factores en el SPQ y un modelo de cuatro dimensiones independientes en el CEQ. En el caso de las evidencias de consistencia interna, se esperaba que los instrumentos presentaran estadísticos por sobre el valor de corte 0,7 propuesto en la literatura. Finalmente, en el caso específico de las evidencias de criterio, se hipotetizó que los instrumentos presentan correlaciones teóricamente plausibles. A continuación se resume esta información y en el apartado Análisis de Datos se argumentan y justifican estas decisiones.

Tabla 3
Hipótesis

Instrumento	Tipo de evidencia	Prueba	Evidencia esperada
SPQ	Contenido	Porcentaje de acuerdo entre evaluadores	Sobre 70%
	Estructura interna	Análisis factorial confirmatorio*	Modelo de dos escalas: Aprendizaje Superficial y Aprendizaje Profundo, con ajustes
	Consistencia Interna	Coficiente alfa de Cronbach	$\alpha \geq 0,7$
	Consistencia Interna	Coficiente omega de McDonald	$\omega \geq 0,7$
	Criterio	Correlación SPQ-CEQ	Superficial-Buena Docencia: $r < 0, p < 0,05$ Superficial-Metas y Objetivos Claros: $r < 0, p < 0,05$ Superficial-Evaluación Apropiaada: $r < 0, p < 0,05$ Superficial-Carga de Trabajo: $r < 0, p < 0,05$ Profundo-Buena Docencia: $r > 0, p < 0,05$ Profundo-Metas y Objetivos Claros: $r > 0, p < 0,05$ Profundo-Evaluación Apropiaada: $r > 0, p < 0,05$ Profundo-Carga de Trabajo: $r > 0, p < 0,05$
CEQ	Contenido	Porcentaje de acuerdo entre evaluadores	Sobre 70%
	Estructura interna	Análisis factorial confirmatorio*	Modelo de cuatro escalas: Buena Docencia, Metas y Objetivos Claros, Evaluación Apropiaada y Carga de Trabajo
	Consistencia Interna	Coficiente alfa de Cronbach	$\alpha \geq 0,7$
	Consistencia Interna	Coficiente omega de McDonald	$\omega \geq 0,7$

Nota. La evidencia de criterio se utiliza por igual en ambos instrumentos, por lo que solo se incluyó una vez. Además, esta hipótesis no se basa únicamente en criterios estadísticos, sino en la evidencia que arrojan diferentes estudios (e.g., González et al., 2011).

* En el caso de los análisis factoriales confirmatorios, la evidencia estadística esperada se muestra en la Tabla 5.

Método

Diseño

El presente estudio usó un enfoque cuantitativo y supuso un diseño de encuesta transversal de alcance descriptivo-relacional (León García & Montero García-Celay, 2003). El contexto de la presente investigación

fue un proceso de formación docente, específicamente, el Diplomado en Docencia Universitaria (DDU) de la Universidad de Santiago de Chile.

El DDU tiene como objetivo mejorar las prácticas educativas de los docentes que forman parte de su plantel desde el año 2009. El programa se orienta a promover la reflexión crítica del docente sobre su práctica, junto con el desarrollo de estrategias y herramientas que le permitan orientar su enseñanza con un foco en el aprendizaje de sus estudiantes. Por ello, se decidió evaluar también a los estudiantes de estos profesores. Hasta la fecha en que se realizó la recolección de información, participaron del programa un total de 509 profesores, quienes completaron un total de 160 horas de trabajo presencial.

Participantes

La amplitud de ambos grupos y el marco de la presente investigación, relacionado específicamente con el impacto de la formación docente en los estudiantes, impulsaron la selección de una muestra de alumnos que considerara la participación de los respectivos profesores en el DDU. Así, se seleccionaron dos grupos a través de un muestreo no probabilístico por conveniencia (León García & Montero García-Celay, 2003).

En el primer grupo se seleccionó a 248 docentes en total. De este total, 124 participantes cursaron el DDU, mientras que los otros 124 docentes fueron seleccionados intentando parear a cada docente con otro de su misma facultad, género y relación contractual, pero que cumpliera con la condición de no haber cursado ningún curso del programa de formación.

La elección del segundo grupo se llevó a cabo posteriormente e implicó la selección de 686 estudiantes. En este caso, se le solicitó acceso a los 248 docentes del primer grupo para recoger la información de sus estudiantes, sin esperar que los grupos de estudiantes quedaran equiparados, debido a la baja tasa de respuesta a este llamado por parte de los docentes. Solo 41 profesores de la muestra total accedieron. De estos 41 docentes, 24 habían completado el DDU, mientras que 17 no habían cursado dicho programa. Respecto de los estudiantes, el mayor porcentaje de participantes correspondió a Ingeniería, Tecnologías y Ciencias Médicas, si bien también hubo participantes de las áreas de Administración y Economía, Ciencias, Humanidades y Química y Biología. Un 51,4% de los estudiantes eran hombres y el 48,6%, mujeres. La media de edad fue de 21.99 años ($DE = 3,57$). Asimismo, un buen porcentaje de los estudiantes consideró que el curso que evaluaron es de los más relevantes del semestre y un porcentaje incluso mayor consideró que era un curso relevante para sí mismos.

Esta información se resume en la Tabla 4.

Tabla 4
Características Generales de la Muestra

	Docentes	Estudiantes
N total (Composición de grupos)	248 profesores (148 con diplomado y 148 sin diplomado)	686 (sin grupos)
Promedio de edad	47.89 años ($DE = 13,04$)	21,99 años ($DE = 3,57$)
Sexo masculino (femenino)	56,03% (43,96%)	51,40% (48,60%)
Facultad de procedencia	Ingeniería (30,4%) Tecnologías (15,2%) Ciencias Médicas (13,2%) Humanidades (13,2%) Química y Biología (10,3%) Ciencias (7,8%) Administración y Economía (4,9%) Otras (4,9%)	Ingeniería (29,0%) Tecnologías (23,7%) Ciencias Médicas (14,7%) Administración y Economía (11,7%) Ciencias (8,7%) Humanidades (8,5%) Química y Biología (3,8%)

Instrumentos

Tal como se indicó en los apartados anteriores, si bien la presente investigación utilizó como referencia teórica el modelo 3P, se focalizó en el eje de procesos. Así, se utilizaron dos instrumentos construidos para evaluar elementos de este eje y se evaluaron posibles evidencias de su validez: el SPQ y el CEQ.

Previamente, se adaptaron al contexto chileno los dos instrumentos utilizados en la presente investigación. Mediante la selección de expertos cualificados se aseguró que el proceso de adaptación tuviese en cuenta las diferencias lingüísticas, psicológicas y culturales del contexto, considerando directrices técnicas y metodológicas sugeridas a nivel internacional (Muñiz, Elosua & Hambleton, 2013).

Para ejecutar el proceso de adaptación de los instrumentos, se consideraron las versiones en inglés utilizadas por Webster et al. (2009) y Biggs, Kember y Leung (2001) del SPQ y el CEQ, respectivamente, y se contrastaron con las versiones traducidas y adaptadas al contexto chileno por González et al. (2011).

Para complementar esta aplicación, se agregaron preguntas sobre variables contextuales. Cada instrumento se detalla a continuación.

Study Process Questionnaire (SPQ). El SPQ corresponde a una escala tipo Likert compuesta por 20 reactivos agrupados en dos dimensiones: Aprendizaje Superficial y Aprendizaje Profundo. Ante cada ítem, el participante debe responder respecto de la frecuencia en que se produce la aseveración, situándose en un continuo cuyo rango va desde 1 (*Nunca o casi nunca*) hasta 5 (*Siempre o casi siempre*). Algunos ejemplos de ítems son “Asisto a la mayoría de las clases de este curso con preguntas en mente para las cuales busco respuesta” y “Trabajo duro en este curso porque encuentro que el contenido es interesante”. La interpretación de los resultados se lleva a cabo comparando los puntajes en ambas escalas en una persona o un grupo determinado, y no existen normas generales como una forma de argumentar a favor de la naturaleza contextual del constructo (Prosser & Trigwell, 2006).

Course Experience Questionnaire (CEQ). Corresponde a un conjunto de escalas tipo Likert que funcionan como indicadores de la calidad de la docencia desde la perspectiva de los estudiantes. Para la presente investigación se consideró pertinente utilizar solamente cuatro escalas que comprenden un total de 17 ítems, las que se ajustaban a la diversidad de los cursos evaluados. Estas fueron Buena Docencia (e.g., “El profesor se esfuerza para que la materia sea interesante”), Metas y Objetivos Claros (e.g., “Usualmente he tenido una idea clara de qué se espera de mí en este curso”), Evaluación Apropiaada (e.g., “Solo se necesita buena memoria para que a uno le vaya bien en este curso”) y Carga de Trabajo (e.g., “En este curso se otorga suficiente tiempo para comprender los contenidos propuestos”). Ante cada ítem, el participante responde respecto del grado de acuerdo con la aseveración, situándose en un continuo cuyo rango va desde 1 (*Muy en desacuerdo*) hasta 5 (*Muy de acuerdo*). El CEQ ha sido utilizado como una herramienta sumativa para evaluar diferentes programas de estudio, si bien en la presente investigación se utilizó una versión construida para evaluar una asignatura específica.

Cuestionario de variables contextuales. En el cuestionario final se incluyeron variables contextuales a través de preguntas elaboradas por el equipo de investigadores. Estas preguntas indagaron elementos específicos de los estudiantes. Se les preguntó por su facultad de procedencia, sexo, edad y relevancia del curso. Estas variables permitieron caracterizar al grupo.

Procedimiento

Los cuestionarios fueron administrados a los estudiantes en la sala de clases durante una única sesión de aproximadamente 30 minutos. En primer lugar se aplicó el cuestionario de variables contextuales, en segundo lugar, el SPQ y finalmente, el CEQ. La aplicación se realizó en forma individual y autoadministrada, en base a un protocolo estandarizado y con un investigador presente. Se administró de forma previa un consentimiento informado, que fue leído a los participantes en conjunto con los objetivos de la investigación, y posteriormente firmado por ellos, en el que se recalcó que la participación era voluntaria y anónima.

Análisis de Datos

Evidencias sobre el contenido de los instrumentos. De forma previa a la administración final de los instrumentos, tanto el SPQ como el CEQ fueron analizados por un grupo de seis especialistas en docencia

universitaria, incluyendo a doctores, máster y estudiantes de doctorado en educación, entre otros. El equipo de especialistas analizó los instrumentos adaptados a partir del modelo *Question Appraisal System/QAS-99* de Willis y Lessler (1999), que permite calcular el grado de acuerdo entre diferentes evaluadores respecto de la pertinencia de los contenidos incluidos en cada instrumento. Más específicamente, este cálculo se desarrolla a través del índice Kappa (Fleiss, Cohen & Everitt, 1969). Para aplicar este sistema, se utilizó el software e-Encuesta.

Evidencias sobre la estructura interna de los instrumentos. Si bien el modelo 3P constituye una herramienta analítica que integra diferentes elementos presentes en los fenómenos educativos, y que tienen una estructura jerárquica, usualmente caracterizada en torno a conglomerados de estudiantes, cursos e institución (Crislip & Heck, 2001), los instrumentos trabajados en este estudio han sido construidos y desarrollados por separado para evaluar diferentes variables. Por esto, en el presente artículo una parte de los análisis se realizaron considerando esta independencia en su utilización y evaluación, mientras en otra parte se analizaron las relaciones entre las puntuaciones obtenidas para dar cuenta del marco referencial en que se insertan.

Los análisis estadísticos fueron realizados con STATA/SE v.12. Primero, se llevó a cabo una caracterización psicométrica de los dos instrumentos para describir su estructura factorial y evaluar el ajuste de distintos modelos teóricos. En la misma línea, también se evaluó su consistencia interna. Finalmente, se calcularon correlaciones de Spearman entre las puntuaciones de los instrumentos para evaluar en qué medida los instrumentos se adecuan al modelo teórico del que se desprenden.

Para describir la estructura factorial de los instrumentos y la comparación de ajuste de distintos modelos teóricos, se realizó un análisis factorial confirmatorio (AFC). Esta elección se debió a que este permite contrastar una teoría específica con las respuestas a determinados indicadores empíricos, lo que establece una conexión entre los argumentos teóricos de determinado marco conceptual y las consecuencias concretas del uso de las puntuaciones de una prueba que se estructura de acuerdo a dicho marco (Batista-Foguet, Coenders & Alonso, 2004). Se usó el método de estimación de mínimos cuadrados no ponderados en cada uno de los modelos contrastados (Manzano Patiño & Zamora Muñoz, 2009). Los criterios de una buena bondad de ajuste de los modelos son resumidos en la Tabla 5.

Tabla 5
Criterios de Bondad de Ajuste Para Modelos Estructurales

Índice	Abreviatura	Criterio
Chi-cuadrado	χ^2	$p > 0,05$
Índice comparativo de bondad de ajuste	CFI	$\geq 0,95$
Índice de Tucker-Lewis	TLI	$\geq 0,95$
Raíz del residuo cuadrático medio estandarizado	SRMR	$< 0,05$

Nota. La raíz del error cuadrático medio de aproximación (RMSEA) no se calculó, debido al bajo tamaño muestral, tal como sugieren Kenny, Kaniskan y McCoach (2015).

Evidencias sobre la consistencia interna de los instrumentos. Para evaluar la consistencia interna de los instrumentos se calcularon dos indicadores: el coeficiente alfa de Cronbach y el coeficiente omega de McDonald (Elosua Oliden & Zumbo, 2008). El primero se calculó por ser el índice más usado en ciencias sociales (Zumbo & Rupp, 2004), lo que permite tener puntos de comparación con otros estudios. El segundo se utilizó porque constituye una de las más recientes alternativas para estimar la consistencia interna, en la medida que considera la naturaleza ordinal de los datos y es más preciso en las estimaciones (Zinbarg, Revelle, Yovel & Li, 2005).

Evidencias sobre la relación de los instrumentos con otros criterios. Si bien se tuvo en consideración que tanto el SPQ como el CEQ se estaban analizando en el presente trabajo y, por tanto, no eran estrictamente criterios entre sí mismos, la literatura muestra relación entre ambos (Ginns & Ellis, 2009; Ramsden, 2003) y, además, se cuenta con evidencias previas de su correlaciones en el contexto chileno

(González et al., 2011). Debido a esto, se buscó detectar la validez de ambos instrumentos a partir del cálculo de las relaciones entre las escalas, mediante el coeficiente de correlación de Spearman. Se utilizó este coeficiente, pues se asumió la naturaleza ordinal de los puntajes de las escalas (Martínez Ortega, Tuya Pendás, Martínez Ortega, Pérez Abreu & Cánovas, 2009). En particular, se analizaron las relaciones internas en cada instrumento y entre los instrumentos, siguiendo las hipótesis planteadas en la Tabla 3.

Resultados

Contenido de los Instrumentos

En el caso de la evidencia sobre el contenido del SPQ y el CEQ, los integrantes del equipo de especialistas que revisó los instrumentos obtuvieron un porcentaje de acuerdo en la aprobación de los ítems de un 84%. Además, se incluyó un apartado de observaciones, donde cada integrante del equipo pudo anotar sus comentarios y sugerencias. A partir de las observaciones de los especialistas a cada uno de los ítems, se incluyeron cambios específicos en los reactivos para que pudiesen ser aplicados sin problemas en el contexto de la Universidad de Santiago.

Estructura Interna de los Instrumentos

Previo a cada uno de los análisis, se llevó a cabo una exploración preliminar de los datos, con el objetivo de evaluar la presencia de casos perdidos, casos atípicos, linealidad y multicolinealidad. Se reportó menos de un 1% de datos perdidos, por lo que no hubo necesidad de imputar información (Muñoz Rosas & Álvarez Verdejo, 2009), y se encontró que varios ítems tenían distribuciones asimétricas, por lo que se usó el método de Satorra-Bentler (Satorra & Bentler, 2001) para corregir los AFC posteriores en la presencia de datos no-normales.

Luego, se evaluó la pertinencia de usar el AFC sobre la matriz de datos del SPQ. Tanto el cálculo del estadístico KMO (0,815) como la prueba de esfericidad de Bartlett (1597,3, $p < 0,001$) sugirieron la adecuación de los datos, por lo que posteriormente se ejecutaron los AFC.

Se propusieron dos modelos. En el primer modelo se especificaron cuatro factores, mientras que en el segundo modelo, dos. Los resultados indicaron que el modelo de cuatro factores no se ajustaba de manera apropiada a los datos: $\chi^2(2547, N = 589) = 13540,60$, $p = 0,043$, CFI = 0,930, TLI = 0,912, SRMR = 0,093. A la par, el modelo de dos factores propuesto por la literatura presentaba un mejor ajuste: $\chi^2(1821, N = 575) = 1450,89$, $p = 0,086$, CFI = 0,965, TLI = 0,939, SRMR = 0,041. El modelo de dos factores se consideró, por tanto, más apropiado para describir los datos observados y se puede analizar en detalle en la Figura 2.

En el caso del CEQ, se llevaron a cabo AFC para cada una de sus escalas. Estas fueron consideradas independientes, puesto que, a diferencia del SPQ, responden a constructos teóricamente no relacionados estructuralmente entre sí (Richardson, 2005). Las pruebas de adecuación del AFC arrojaron resultados positivos (KMO = 0,888; prueba de esfericidad de Bartlett = 4780,88, $p < 0,001$). Los índices de ajuste indicaron un buen ajuste para la escalas Buena Docencia, $\chi^2(5,36, N = 668) = 100,38$, $p = 0,084$, CFI = 0,997, TLI = 0,985, SRMR = 0,011; Metas y Objetivos Claros, $\chi^2(483,4, N = 657) = 855,84$, $p = 0,061$, CFI = 0,975, TLI = 0,951, SRMR = 0,038; Carga de Trabajo, $\chi^2(586,3, N = 569) = 980,31$, $p = 0,075$, CFI = 0,971, TLI = 0,949, SRMR = 0,039, y Evaluación Apropriada, $\chi^2(1994, N = 579) = 522,72$, $p = 0,053$, CFI = 0,985, TLI = 0,956, SRMR = 0,024.

El diagrama estructural se observa en la Figura 2.

Consistencia Interna de los Instrumentos

Los resultados se exponen en la Tabla 6, que resume como todas las escalas obtuvieron índices adecuados de consistencia interna (Elosua Oliden & Zumbo, 2008).

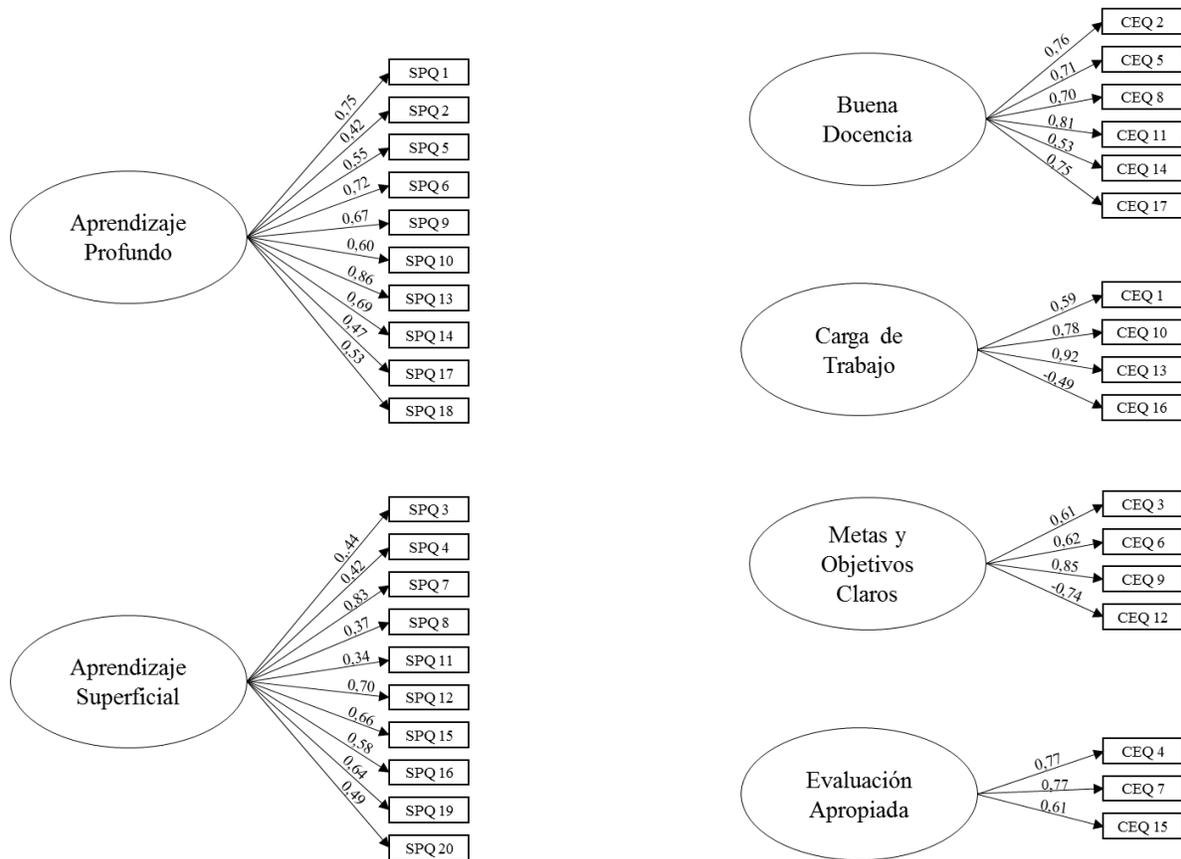


Figura 2. Modelos estructurales obtenidos. De izquierda a derecha: SPQ y CEQ. Los valores en las líneas representan correlaciones. Asimismo, no se muestran los errores de medida, si bien se consideraron.

Tabla 6
Consistencia Interna de los Instrumentos

Instrumento	Escala	Número de Ítems	Coefficiente alfa de Cronbach	Coefficiente omega de McDonald
SPQ	Enfoque Profundo	10	0,8475	0,8585
	Enfoque Superficial	10	0,8040	0,8160
CEQ	Buena Docencia	6	0,8585	0,8675
	Metas y Objetivos Claros	4	0,7408	0,7419
	Carga de Trabajo Apropiciada	4	0,7817	0,7907
	Evaluación Apropiciada	3	0,6977	0,7097

Relación de los Instrumentos Dentro y Entre Sí

Se observó una serie de correlaciones al interior de los instrumentos. En primer lugar, las escalas del SPQ se correlacionaron entre sí de manera indirecta. En el caso de las escalas del CEQ, las relaciones fueron tanto directas como indirectas y fluctuaron entre niveles bajos y moderados.

En segundo lugar, es interesante observar que se encontraron correlaciones entre las escalas del SPQ y CEQ. Así, existen relaciones moderadas y directas entre Aprendizaje Profundo y Buena Docencia y entre Aprendizaje Profundo y Metas y Objetivos Claros. También se hallaron relaciones bajas y directas entre

Aprendizaje Profundo y Evaluación Apropriadada y entre Aprendizaje Profundo y Carga de Trabajo Apropriadada, relaciones bajas e inversas entre Aprendizaje Superficial y Buena Docencia, entre Aprendizaje Superficial y Metas y Objetivos Claros y entre Aprendizaje Superficial y Carga de Trabajo Apropriadada. Finalmente, también se reportó una relación inversa y moderada entre Aprendizaje Superficial y Evaluación Apropriadada. Los resultados se pueden observar en la Tabla 7.

Tabla 7
Correlaciones Dentro y Entre el SPQ y el CEQ

Factor	SPQ Aprendizaje Profundo	SPQ Aprendizaje Superficial	CEQ Buena Docencia	CEQ Metas y Objetivos Claros	CEQ Evaluación Apropriadada	CEQ Carga de Trabajo Apropriadada
SPQ Aprendizaje Profundo	1					
SPQ Aprendizaje Superficial	-0,617**	1				
CEQ Buena Docencia	0,502**	-0,325**	1			
CEQ Metas y Objetivos Claros	0,439**	-0,302**	0,642**	1		
CEQ Evaluación Apropriadada	0,203**	-0,513**	0,264**	0,268**	1	
CEQ Carga de Trabajo Apropriadada	0,147**	-0,089*	0,465**	0,368**	0,081*	1

Nota. $N = 635$. Se consideró el menor N común para todos los cruces de variables representados en la tabla.

** La correlación es significativa al nivel 0,01.

* La correlación es significativa al nivel 0,05.

Discusión y Conclusiones

La presente investigación tuvo como objetivo adaptar y aportar nuevas evidencias de validez del uso de los cuestionarios SPQ y CEQ en el análisis de las iniciativas de formación docente en la educación superior chilena.

El primer tema relevante que se consideró fue la adaptación apropiada de los instrumentos. Las adaptaciones realizadas en el marco de este estudio fueron conducidas siguiendo estándares sugeridos por la literatura especializada (Muñiz et al., 2013) y permitieron superar las debilidades de los estudios previos (González et al., 2011, 2012; Montenegro Maggio & González Ugalde, 2013), en la medida en que consideraron las diferencias lingüísticas, psicológicas y culturales específicas del contexto. El resultado de este proceso son instrumentos adaptados al menos preliminarmente a las características del contexto universitario.

El segundo tema que se trabajó fue la exploración de la utilidad de los instrumentos en el contexto chileno de la formación docente. En esta línea, se analizaron cuatro tipos de evidencias. Resulta interesante comprobar que en el caso del primer tipo de evidencias hubo acuerdos en el equipo evaluador en la adecuación de los reactivos para la evaluación de programas de formación docente. Esto permitió generar instrumentos más adecuados para la realidad institucional específica del contexto del presente estudio y, además, introducir algunos cambios que aportaron para que la evidencia fuera más consistente (Willis & Lessler, 1999). Asimismo, es una evidencia del posible uso de los instrumentos para la evaluación de programas de formación docente en el país y, específicamente, en la Universidad de Santiago.

Sobre el segundo conjunto de evidencias, en el caso del SPQ, las evidencias apoyan su uso como un instrumento de dos factores. En el caso del CEQ, los resultados de las cuatro escalas analizadas aportan evidencias sobre la pertinencia del instrumento para evaluar la percepción de los estudiantes respecto de la calidad de la enseñanza en sus cursos a través de cuatro escalas independientes. Toda esta información aporta evidencias a favor de las hipótesis propuestas.

Más específicamente sobre el SPQ, las evidencias estructurales no apuntaron hacia la hipótesis de una estructura tetradimensional sino, por el contrario, apuntaron hacia una estructura de dos dimensiones. Si bien diferentes estudios internacionales han reportado estructuras de cuatro factores, que diferencian entre los motivos y las estrategias que siguen los alumnos para abordar sus aprendizajes (Fryer et al., 2012; Stes et al., 2013), la evidencia obtenida con los datos de la presente investigación más bien concuerda con los hallazgos preliminares de González et al. (2011), así como con otras evidencias aportadas por estudios

conducidos fuera de Chile (e.g., Munshi et al., 2012; Socha & Sigler, 2014), en los que se propone la utilización del instrumento para identificar si el estudiante ocupa en mayor medida un enfoque de aprendizaje profundo o un aprendizaje superficial.

Sobre el CEQ, las evidencias de independencia en las escalas son similares a las obtenidas en otros contextos (e.g., Fryer et al., 2012). Más en detalle, que las cuatro escalas sean independientes es un resultado reportado en todos los estudios examinados que se señalan en la Tabla 2 y las evidencias para el caso chileno no parecen diferir de estas (González et al., 2011, 2012).

Estas evidencias también se complementan con los resultados del tercer tipo de evidencias, que se desprenden de la consistencia interna que se observó en las puntuaciones del SPQ y el CEQ (Elosua Oliden & Zumbo, 2008). Si bien esto no ocurre en el caso específico de la escala de Evaluación Apropiaada, esto puede deberse tanto a la cantidad de ítems o a características de la muestra. Este argumento requiere posteriores desarrollos.

En cuarto lugar, complementariamente se halló que un enfoque profundo de aprendizaje se relaciona con una percepción de buena enseñanza y de metas y objetivos claros, mientras que un enfoque superficial se relaciona con una percepción de incoherencia en las evaluaciones y con una alta carga de trabajo. Si bien la relación es baja, esta evidencia parece especialmente interesante en tanto es coherente con los hallazgos de Prosser y Trigwell (2001) y González et al. (2011) y, a su vez, reflejan una consistencia teórica que apoya la validez del uso de los instrumentos desarrollados en el contexto de la formación docente. En otras palabras, tanto las estructuras de los instrumentos adaptados al contexto chileno como la relación entre los criterios concuerdan con la descripción teórica propuesta por el modelo 3P.

Sin embargo, es importante mencionar que, de acuerdo a los planteamientos actuales relacionados al concepto de validez, los argumentos presentados son insuficientes para sustentar todos los posibles usos de los instrumentos (Kane, 2013). Si bien las conclusiones de la presente investigación apuntan de manera preliminar a que los instrumentos presentan evidencias de consistencia interna y validez en el contexto de la investigación y evaluación de programas de formación docente, futuras investigaciones podrían continuar analizando otro tipo de evidencia, como aquella que se relaciona con los procesos de respuesta involucrados en los instrumentos y las consecuencias que podrían derivar del uso de estos instrumentos (Kane, 2013).

En el presente estudio emergen cuatro limitaciones. La primera tiene que ver con la estrategia de análisis. La estructura jerárquica y multinivel de los fenómenos educativos requiere análisis complejos que consideren la naturaleza anidada de la información (Crislip & Heck, 2001). Esto no se desarrolló, debido a la decisión de considerar únicamente los instrumentos para estudiantes por separado, como han hecho otros estudios previos (Montenegro Maggio & González Ugalde, 2013; Munshi et al., 2012; Trigwell & Prosser, 2004). La segunda limitación tiene que ver con el tipo de muestra que se incluyó en este estudio. El hecho de que fuera por conveniencia no permite generalizar los resultados. Además, al tratarse de alumnos que cumplen con una serie de características específicas, como, por ejemplo, haber respondido al llamado o haber sido seleccionados por su profesor, se abre la posibilidad de que los resultados puedan estar influenciados por otros motivos que no se consideraron. La tercera limitación se relaciona con la ausencia de variables que evaluarán los demás ejes del modelo 3P. Si bien no se consideraron, ya que el objetivo central era adaptar y evaluar instrumentos del eje de procesos, hubiese sido interesante integrar variables de los demás ejes para dar cuenta de las interrelaciones que se producen en los diferentes momentos del modelo 3P, especialmente con los resultados de aprendizaje y las variables de presagio de los estudiantes. En este último caso, es necesario mencionar que no fue posible controlar el efecto de otras características de los estudiantes (e.g., nivel socioeconómico o experiencias previas), debido a la poca dispersión de las respuestas de los cuestionarios. Por ello, si bien los resultados pueden ser explicados por las dificultades metodológicas, también pudiesen estar influidos por estas características y, ante ello, es indispensable tomarlos con cautela.

Finalmente, este estudio constituye no solo un aporte a la investigación en educación, sino que el uso de los instrumentos examinados puede contribuir a la evaluación de programas de formación docente y a la discusión de políticas públicas enfocadas al mejoramiento de la calidad en la educación superior chilena.

Referencias

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Batista-Foguet, J. M., Coenders, G. & Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina Clínica*, 122(Suplemento 1), 21-27. doi:10.1157/13057542

- Biggs, J. B. (1987). *Student approaches to learning and studying. Research monograph*. Melbourne, Australia: Australian Council for Educational Research.
- Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research & Development*, 18, 57-75. doi:10.1080/0729436990180105
- Biggs, J. (2008). *Calidad del aprendizaje universitario*. Madrid, España: Narcea.
- Biggs, J., Kember, D. & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71, 133-149. doi:10.1348/000709901158433
- Byrne, M. & Flood, B. (2003). Assessing the teaching quality of accounting programmes: An evaluation of the Course Experience Questionnaire. *Assessment & Evaluation in Higher Education*, 28, 135-145. doi:10.1080/02602930301668
- Cid-Sabucedo, A., Pérez-Abellás, A. & Zabalza, M. A. (2009). Las prácticas de enseñanza declaradas de los “mejores profesores” de la Universidad de Vigo. *Relieve*, 15(2), 1-29.
- Coll, C. & Sánchez, E. (2008). Presentación. El análisis de la interacción alumno-profesor: líneas de investigación. *Revista de Educación de España*, 346, 15-32.
- Coll, C. & Solé, I. (2001). Enseñar y aprender en el contexto del aula. En C. Coll, J. Palacios & A. Marchesi (Comps.), *Desarrollo psicológico y educación 2: Psicología de la educación* (pp. 59-81). Madrid, España: Alianza.
- Crislip, M. A. & Heck, R. H. (2001, Abril). *Testing a multilevel model*. Ponencia presentada en la 2001 AERA Annual Meeting, American Educational Research Association, Seattle, WA, Estados Unidos.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334. doi:10.1007/BF02310555
- De la Fuente Arias, J., Martínez Vicente, J. M., Peralta Sánchez, F. J. & Berben, A. (2010). Percepción del proceso de enseñanza-aprendizaje y rendimiento académico en diferentes contextos instruccionales de la educación superior. *Psicothema*, 22, 806-812.
- Doménech Betoret, F., Jara Jiménez, P. & Rosel Remírez, J. (2004). Percepción del proceso de enseñanza/aprendizaje desarrollado en Psicoestadística I y su incidencia en el rendimiento. *Psicothema*, 16, 32-38.
- Dunkin, M. J. & Biddle, B. J. (1974). *The study of teaching*. Oxford, Reino Unido: Holt, Rinehart & Winston.
- Elosua Oliden, P. & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20, 896-901.
- Entwistle, N. (1997). Reconstituting approaches to learning: A response to Webb. *Higher Education*, 33, 213-218. doi:10.1023/A:1002930608372
- Fleiss, J. L., Cohen, J. & Everitt, B. S. (1969). Large sample standard errors of Kappa and weighted Kappa. *Psychological Bulletin*, 72, 323-327. doi:10.1037/h0028106
- Fryer, L. K., Ginns, P., Walker, R. A. & Nakao, K. (2012). The adaptation and validation of the CEQ and the R-SPQ-2F to the Japanese tertiary environment. *British Journal of Educational Psychology*, 82, 549-563. doi:10.1111/j.2044-8279.2011.02045.x
- Gibbs, G. & Coffey, M. (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education*, 5, 87-100. doi:10.1177/1469787404040463
- Ginns, P. & Ellis, R. A. (2009). Evaluating the quality of e-learning at the degree level in the student experience of blended learning. *British Journal of Educational Technology*, 40, 652-663. doi:10.1111/j.1467-8535.2008.00861.x
- González, C., Montenegro, H. & López, L. (2012). Análisis de confiabilidad y de validez del instrumento Course Experience Questionnaire (CEQ). *Educación y Educadores*, 15, 63-78. doi:10.5294/edu.2012.15.1.4
- González, C., Montenegro, H., López, L., Munita, I. & Collao, P. (2011). Relación entre la experiencia de aprendizaje de estudiantes universitarios y la docencia de sus profesores. *Calidad en la Educación*, 35, 21-49. doi:10.4067/S0718-45652011000200002
- Hogan, T. P. (2003/2004). *Pruebas psicológicas: una introducción práctica* (J. Cisneros Herrera, Trad.; Título original: Psychological testing: A practical introduction). México DF, México: Manual Moderno.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73. doi:10.1111/jedm.12000
- Kenny, D. A., Kaniskan, B. & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44, 486-507. doi:10.1177/0049124114543236
- Lawless, C. J. & Richardson, J. T. E. (2002). Approaches to studying and perceptions of academic quality in distance education. *Higher Education*, 44, 257-282. doi:10.1023/A:1016315114558
- León García, O. G. & Montero García-Celay, I. (2003). *Métodos de investigación en psicología y educación* (3ª ed.). Madrid, España: McGraw Hill.
- Longden, B. (2010). Rowena Murray (ed): The scholarship of teaching and learning in higher education. *Higher Education*, 59, 799-800. doi:10.1007/s10734-009-9267-0
- Manzano Patiño, A. & Zamora Muñoz, S. (2009). *Sistema de ecuaciones estructurales: una herramienta de investigación* (Cuaderno Técnico 4). México DF, México: Centro Nacional de Evaluación para la Educación Superior.
- Martínez Ortega, R. M., Tuya Pendás, L. C., Martínez Ortega, M., Pérez Abreu, A. & Cánovas, A. M. (2009). El coeficiente de correlación de los rangos de Spearman: caracterización. *Revista Habanera de Ciencias Médicas*, 8(2), 1-19.
- Marton, F., Hounsell, D. & Entwistle, N. (Eds.) (1984). *The experience of learning*. Edinburgh, Reino Unido: Scottish Academic Press.
- Montenegro Maggio, H. & González Ugalde, C. (2013). Análisis factorial confirmatorio del cuestionario “Enfoques de Docencia Universitaria” (Approaches to Teaching Inventory, ATI-R). *Estudios Pedagógicos*, 39(2), 213-230. doi:10.4067/S0718-07052013000200014
- Munshi, F. M., Al-Rukban, M. O. & Al-Hoqail, I. (2012). Reliability and validity of an Arabic version of the Revised Two-Factor Study Process Questionnaire R-SPQ-2F. *Journal of Family and Community Medicine*, 19, 33-37. doi:10.4103/2230-8229.94010
- Muñiz, J., Elosua, P. & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157.
- Muñoz Rosas, J. F. & Álvarez Verdejo, E. (2009). Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Plus. *Revista de Métodos Cuantitativos Para la Economía y la Empresa*, 7, 3-30.
- Pichardo Martínez, M. C., García Berbén, A. B., De la Fuente Arias, J. & Justicia Justicia, F. (2007). El estudio de las expectativas en la universidad: análisis de trabajos empíricos y futuras líneas de investigación. *Revista Electrónica de Investigación Educativa*, 9(1), 1-15.
- Postareff, L. & Lindblom-Ylänne, S. (2008). Variation in teachers' descriptions of teaching: Broadening the understanding of teaching in higher education. *Learning and Instruction*, 18, 109-120. doi:10.1016/j.learninstruc.2007.01.008

- Postareff, L., Lindblom-Ylänne, S. & Nevgi, A. (2007). The effect of pedagogical training on teaching in higher education. *Teaching and Teacher Education*, 23, 557-571. doi:10.1016/j.tate.2006.11.013
- Prosser, M. & Trigwell, K. (2001). *Understanding learning and teaching: The experience in higher education*. Buckingham, Reino Unido: Society for Research Into Higher Education/Open University Press.
- Prosser, M. & Trigwell, K. (2006). Confirmatory factor analysis of the Approaches to Teaching Inventory. *British Journal of Educational Psychology*, 76, 405-419. doi:10.1348/000709905X43571
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: the Course Experience Questionnaire. *Studies in Higher Education*, 16, 129-150. doi:10.1080/03075079112331382944
- Ramsden, P. (2003). *Learning to teach in higher education*. Abingdon, Reino Unido: Routledge.
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30, 387-415. doi:10.1080/02602930500099193
- Ruiz, F. A. & Schumacher, C. (2008). Evaluación del aprendizaje universitario. *Educación y Educadores*, 11(2), 91-105.
- Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514. doi:10.1007/BF02296192
- Shulman, L. S. (1986/1989). Paradigmas y programas de investigación en el estudio de la enseñanza: una perspectiva contemporánea (O. Castillo, Trad.; Título original: Paradigms and research programs in the study of teaching: A contemporary perspective). En M. C. Wittrock (Ed.), *La investigación de la enseñanza I: Enfoques, teorías y métodos* (pp. 9-91) (O. Castillo & G. Vitale, Trad.; Título original: Handbook of research on teaching, caps. 1, 2 y 8). Barcelona, España: Paidós.
- Socha, A. & Sigler, E. A. (2014). Exploring and "reconciling" the factor structure for the Revised Two-factor Study Process Questionnaire. *Learning and Individual Differences*, 31, 43-50. doi:10.1016/j.lindif.2013.12.010
- Spooren, P., Brockx, B. & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83, 598-642. doi:10.3102/0034654313496870
- Stes, A., De Maeyer, S., Gijbels, D. & Van Petegem, P. (2012). Instructional development for teachers in higher education: Effects on students' perceptions of the teaching-learning environment. *British Journal of Educational Psychology*, 82, 398-419. doi:10.1111/j.2044-8279.2011.02032.x
- Stes, A., De Maeyer, S. & Van Petegem, P. (2013). Examining the cross-cultural sensitivity of the Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) and validation of a Dutch version. *Plos One*, 8(1), 1-7. doi:10.1371/journal.pone.0054099
- Trigwell, K. & Prosser, M. (2004). Development and use of the Approaches to Teaching Inventory. *Educational Psychology Review*, 16, 409-424. doi:10.1007/s10648-004-0007-9
- Trigwell, K., Prosser, M. & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57-70. doi:10.1023/A:1003548313194
- Webster, B. J., Chan, W. S. C., Prosser, M. T. & Watkins, D. A. (2009). Undergraduates' learning experience and learning process: Quantitative evidence from the East. *Higher Education*, 58, 375-386. doi:10.1007/s10734-009-9200-6
- Willis, G. B. & Lessler, J. T. (1999). *Question Appraisal System QAS-99*. Rockville, MD: Research Triangle Institute.
- Wilson, K. L., Lizzio, A. & Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education*, 22, 33-53. doi:10.1080/03075079712331381121
- Zabalza, M. A. (2003). *Competencias docentes del profesorado universitario: calidad y desarrollo profesional*. Madrid, España: Narcea.
- Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123-133. doi:10.1007/s11336-003-0974-7
- Zumbo, B. D. & Rupp, A. A. (2004). Responsible modelling of measurement data for appropriate inferences: Important advances in reliability and validity theory. En D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 73-92). Thousand Oaks, CA: SAGE.

Anexo

Cuestionarios
Procesos de Aprendizaje (SPQ)
Experiencias de Aprendizaje (CEQ)

SECCIÓN I: Procesos de Aprendizaje (SPQ)

A continuación se presenta un conjunto de afirmaciones que contemplan diferentes alternativas de respuesta. Para contestar, considera que **LAS PREGUNTAS SE REFIEREN A ESTE CURSO EN ESPECÍFICO**.

Marca con una cruz la alternativa que expresa de manera más precisa tu opinión respecto de cada afirmación:	Muy en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Muy de acuerdo
(1) En ocasiones, estudiar para este curso me proporciona un sentimiento de profunda satisfacción personal.					
(2) Cuando estudio un tema de este curso, no estoy conforme hasta que le dedico el tiempo suficiente para formar mis propias conclusiones.					
(3) Mi propósito es aprobar este curso haciendo el menor trabajo posible.					
(4) Solo estudio seriamente la materia que se pasa en clases o que está en el programa de este curso.					
(5) Siento que casi cualquier tema de este curso podría ser interesante, si le dedico tiempo a trabajar en él.					
(6) Frecuentemente paso tiempo extra buscando información adicional sobre los temas de este curso, ya que me parecen interesantes.					
(7) No encuentro este curso muy interesante, así que trato de hacer el mínimo trabajo.					
(8) Aprendo algunas cosas de este curso repasándolas una y otra vez hasta que las sé de memoria, aunque no las comprenda.					
(9) Creo que estudiar temas de este curso puede ser en ocasiones tan emocionante como una buena novela o película.					
(10) Autoevalúo mis conocimientos sobre temas relevantes de este curso, hasta que los entiendo por completo.					
(11) Puedo aprobar la mayoría de las evaluaciones de este curso memorizando los temas más relevantes, más que tratando de comprenderlos a fondo.					
(12) En este curso generalmente me limito a estudiar solo lo que establece el profesor, porque creo que es innecesario hacer trabajo extra.					
Marca con una cruz la alternativa que expresa de manera más precisa tu opinión respecto de cada afirmación:	Muy en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Muy de acuerdo
(13) Trabajo duro en este curso porque encuentro que el contenido es interesante.					
(14) Dedico una buena parte de mi tiempo libre a buscar más información sobre temas interesantes que hemos revisado en este curso.					
(15) Creo que en este curso es poco útil estudiar los temas en profundidad; eso confunde y hace perder el tiempo, cuando lo único que se necesita es un conocimiento general de los temas.					
(16) El(La) profesor(a) de este curso no debería esperar que los estudiantes dediquen mucho tiempo a estudiar contenidos que se sabe que no serán considerados en las evaluaciones.					
(17) Asisto a la mayoría de las clases de este curso con preguntas en mente para las cuales busco respuesta.					
(18) Me esfuerzo por revisar la mayor parte del material de estudio recomendado para las para las clases de este curso.					
(19) Para mí tiene poco sentido estudiar el material que probablemente no será considerado en las evaluaciones de este curso.					
(20) La mejor forma de aprobar este curso es tratar de memorizar las respuestas a preguntas que probablemente entrarán en las evaluaciones.					

SECCIÓN II: Experiencias de Aprendizaje (CEQ)

A continuación se presenta un conjunto de afirmaciones que contemplan diferentes alternativas de respuesta. Para contestar, considera que **LAS PREGUNTAS SE REFIEREN A ESTE CURSO EN ESPECÍFICO**.

Marca con una **X** la alternativa que expresa de manera más precisa tu opinión respecto de cada afirmación:

	Muy en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Muy de acuerdo
(21) La cantidad de trabajo asociado a este curso impide que toda la materia se pueda entender en profundidad.					
(22) El(La) profesor(a) de este curso me motiva a hacer mi mejor trabajo.					
(23) Es fácil conocer las exigencias del trabajo esperado en este curso.					
(24) Solo se necesita buena memoria para que a uno le vaya bien en este curso.					
(25) El(La) profesor(a) se esfuerza para que la materia sea interesante.					
(26) El(La) profesor(a) dejó en claro desde el inicio del curso lo que se esperaba de los estudiantes.					
(27) El(La) profesor(a) parece más interesado en evaluar lo que he memorizado que lo que he comprendido en el curso.					
(28) El(La) profesor(a) es muy bueno(a) explicando la materia.					
(29) Usualmente he tenido una idea clara de qué se espera de mí en este curso.					
(30) Siento mucha presión como estudiante de este curso.					
(31) El(La) profesor(a) me da retroalimentación sobre mi progreso en el curso.					
(32) Generalmente ha sido difícil descubrir qué se espera de mí en este curso.					
(33) La carga de trabajo en el curso es muy pesada.					
(34) El(La) profesor(a) dedica mucho tiempo para entregarme comentarios sobre mi desempeño en el curso.					
(35) En las evaluaciones el(la) profesor(a) solo hace preguntas sobre información textual de la materia.					
(36) En este curso se otorga suficiente tiempo para comprender los contenidos propuestos.					
(37) El(La) profesor(a) se esfuerza por entender las dificultades que pudiera estar teniendo con mi desempeño en el curso.					

Fecha de recepción: Marzo de 2015.

Fecha de aceptación: Agosto de 2016.