



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

FACULTAD DE EDUCACIÓN

DEPARTAMENTO DE CURRÍCULUM, TECNOLOGÍA Y EVALUACIÓN

CONSTRUCCIÓN Y VALIDACIÓN DEL
INSTRUMENTO DE MEDICIÓN DE INTENCIONES CONDUCTUALES DEL
PROFESORADO HACIA LA EVALUACIÓN DE APRENDIZAJES:
UNA APLICACIÓN DE LA TEORÍA DE CONDUCTA PLANIFICADA

POR

GABRIELA FRANCISCA RODRÍGUEZ MORALES

Proyecto de Intervención presentado a la Facultad de Educación
de la Pontificia Universidad Católica de Chile, para optar al grado de
Magister en Educación con mención en Evaluación de Aprendizajes

Profesora Guía:

CARLA ELVIRA FÖRSTER MARÍN

Noviembre, 2016

Santiago, Chile

TABLA DE CONTENIDOS

INDICE DE TABLAS	5
INDICE DE FIGURAS.....	7
AGRADECIMIENTOS	8
RESUMEN.....	9
ABSTRACT	10
INTRODUCCIÓN	11
I. ANTECEDENTES Y REFERENTES TEÓRICOS	13
1.1. Evaluación de aprendizajes	13
1.1.1. Modelos de la evaluación de aprendizajes.....	14
1.1.1.1. Modelo psicométrico.....	15
1.1.1.2. Modelo sistémico o comunicativo.....	17
1.1.1.3. Modelo comunicativo o psicosocial	21
1.1.2. Antecedentes contextuales de las prácticas evaluativas del profesorado en la actualidad	25
1.2. Medición de variables determinantes del comportamiento.....	28
1.2.1. Medición de actitudes	29
1.2.1.1. Modelo de tres componentes de actitud	30
1.2.2. Medición de la intención conductual	31
1.2.2.1. Teoría de la Acción Razonada.....	31
1.2.2.2. Teoría de Conducta Planificada	33

II. OBJETIVOS.....	36
2.1. Objetivo general	36
2.2. Objetivos específicos	36
III. METODOLOGÍA	37
3.1. Construcción de ítems	37
3.2. Validación de ítems.....	41
3.2.1. Prepiloteo	41
3.2.1.1. Juicio de expertos	41
3.2.1.2. Pretest cognitivo	43
3.2.2. Piloteo	46
3.2.2.1. Participantes	46
3.2.2.2. Características del instrumento	48
3.2.2.3. Plan de análisis	50
3.3. Preparación de instrumento final	54
IV. RESULTADOS	55
4.1. Estadísticos descriptivos	55
4.2. Consistencia interna	55
4.3. Capacidad discriminativa de los ítems	55
4.4. Análisis factorial	56
4.4.1. Análisis factorial del instrumento completo	56
4.4.2. Análisis factorial del instrumento por escalas	60
4.5. Análisis de diferencias entre grupos	66
4.6. Análisis consolidado de reactivos	67

V. DISCUSIÓN	72
VI. LIMITACIONES Y PROYECCIONES DEL CUESTIONARIO.....	77
6.1. Limitaciones del cuestionario	77
6.2. Proyecciones del cuestionario	78
REFERENCIAS BIBLIOGRÁFICAS.....	80

INDICE DE TABLAS

Tabla 1: Teorías de aprendizaje y componentes evaluativos subyacentes a los modelos de evaluación educativa	15
Tabla 2: Requerimientos técnicos de calidad desde la evaluación de aprendizajes a nivel de aula	20
Tabla 3: Categorías consideradas en la construcción de ítems de acuerdo al Modelo comunicativo de evaluación de aprendizajes	38
Tabla 4: Escalas y subescalas consideradas en la construcción de ítems de acuerdo a la Teoría de Conducta Planificada	39
Tabla 5: Distribución de ítems contruidos de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada	40
Tabla 6: Distribución de ítems con porcentaje de aprobación en juicio de expertos igual o superior a 80%, de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada.....	42
Tabla 7: Distribución de ítems seleccionados para pretest cognitivo, de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada	44
Tabla 8: Distribución de ítems seleccionados tras realización de pretest cognitivo, de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada.....	45
Tabla 9: Tabla de especificaciones del instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes, versión pilotaje.....	49
Tabla 10: Análisis psicométricos realizados en el pilotaje y programas estadísticos empleados.....	50

Tabla 11: Varianza explicada basada en los autovalores del Análisis factorial exploratorio con seis factores.....	58
Tabla 12: Caracterización de factores extraídos del Análisis Factorial del instrumento completo.....	59
Tabla 13: Características del Análisis factorial del instrumento por escalas que corroboran la pertinencia del análisis.....	61
Tabla 14: Estadísticos descriptivos multivariados y univariados por escala	61
Tabla 15: Caracterización de factores extraídos del Análisis Factorial, escala Componentes de la intención conductual.....	64
Tabla 16: Caracterización de factores extraídos del Análisis Factorial, escala Conducta	65
Tabla 17: Características del Análisis factorial del instrumento por escalas, versión post pilotaje, que corroboran la pertinencia del análisis.....	68
Tabla 18: Tabla de especificaciones del instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes, versión post pilotaje	70

INDICE DE FIGURAS

Fig. 1: Modelo de tres componentes de actitud.....	30
Fig. 2: Teoría de la Acción Razonada.....	32
Fig. 3: Teoría de Conducta Planificada.....	33
Fig. 4: Porcentaje de participantes que imparte cada asignatura	47
Fig. 5: Curva característica de ítems que discriminan inversamente (29 y 80).	56
Fig. 6: Gráfico de sedimentación de Análisis factorial del instrumento completo con seis factores.	57
Fig. 7: Gráficos de sedimentación por escala.....	62
Fig. 8: Gráfico de sedimentación por escala, versión post pilotaje.....	69

AGRADECIMIENTOS

Este estudio constituye la culminación de un proceso de aprendizaje y desarrollo académico y profesional, el cual no habría sido posible sin el apoyo y comprensión de quienes han confiado en mis capacidades; razón por la cual, quiero expresar mis más sinceros agradecimientos, a todos quienes contribuyeron a la realización de este estudio.

En primer lugar, a mi hija Sol, quien con su llegada fue el principal motor para decidirme a continuar con mi formación profesional; quien ha cedido su tiempo para que yo dedicase largas horas a mis estudios y a mi trabajo, y a quien le debo la fuerza y la inspiración para poder llevar a cabo éste y otros desafíos.

A mi familia, por apoyarme en esta decisión y colaborar en el cuidado de mi hija cuando el estudio o el trabajo así lo requería; por alentarme en momentos de dificultad, por valorar mis esfuerzos y compartir la alegría de mis logros; por enseñarme a no darme por vencida y a enfrentar los retos como una oportunidad de crecimiento.

A Carla Förster, profesora y amiga, quien confió en mis manos el desarrollo de este estudio, proporcionándome consejo y guía en esta experiencia. Asimismo, agradezco también a CONICYT Chile (CONICYT-PCHA/Magíster Nacional/2013-50140196), por financiar mis estudios de postgrado, considerando que a través de ellos podré aportar al desarrollo del país.

A mis profesores, compañeros, amigos, conocidos... a todos quienes durante estos dos años formaron parte importante de mi vida; a quienes no vi tanto y a quienes ya no están, pero habrían estado orgullosos de este paso; a quienes siguen estando, no tan presentes, pero no por ello lejanos.

A todos ellos, gracias.

Este estudio se los debo a ustedes.

RESUMEN

El objetivo de este estudio es construir y validar un instrumento que mida las intenciones conductuales del profesorado hacia la evaluación de aprendizajes.

Inicialmente se construyeron 684 ítems, basados en literatura especializada en evaluación de aprendizajes y en la Teoría de Conducta Planificada. A partir del juicio de expertos y de la realización de un pretest cognitivo, se seleccionaron 80 ítems para una aplicación masiva del cuestionario a una muestra de 509 docentes de Enseñanza Básica y Media del sistema escolar chileno. Se realizaron análisis sobre la capacidad discriminativa de los ítems y de diferencia entre grupos, además de estudiar la consistencia interna y la estructura factorial del instrumento.

Considerando dicha información, se propuso una versión final del cuestionario conformada por 30 ítems y tres escalas: Componentes de la intención conductual (Actitud, Norma subjetiva y Control percibido), Intención conductual y Conducta. El instrumento completo presentó un índice de consistencia interna óptimo (alfa de Cronbach de 0,868). En más de un 75% de los reactivos se obtuvo un índice de capacidad discriminativa sobre 0,30; además, ninguno de los ítems seleccionados presentó diferencias significativas entre grupos de acuerdo al género y a la región.

En consecuencia, se obtuvieron indicadores psicométricos aceptables para el cuestionario completo y para sus dimensiones, logrando con ello evidencia a favor de su validez. De este modo, el cuestionario construido y validado en este estudio, demuestra ser un instrumento robusto para la medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes.

Palabras clave: Construcción de un instrumento, Validación de un instrumento, Teoría de conducta planificada, Evaluación de aprendizajes, Evaluación al interior del aula, Prácticas evaluativas del profesorado.

ABSTRACT

This study aims to construct and validate a measuring instrument of teacher behavioral intentions towards the learning assessment.

Based on a review of learning assessment literature and on the Theory of planned behavior, 684 items were built. From the expert judgment and a cognitive pretest, 80 items were selected and piloted in a massive questionnaire. A total of 509 teachers from primary and secondary education of the Chilean school system were surveyed. Item discrimination, difference between samples, instrument internal consistency and instrument factor structure were studied and analyzed.

Considering such information, a final version of the questionnaire was proposed, which was formed by 30 items and three scales: Components of behavioral intention (Attitude, Subjective norm and Perceived control), Behavioral intention and Behavior. The proposed instrument presented an adequate level of internal consistency (Cronbach's alpha index of 0,868). Over 75% of them have an item discrimination index of more than 0,30. Moreover, the items showed no significant difference between samples by gender and region.

Consequently, acceptable indicators of psychometric indicators were obtained for the full instrument and for its dimensions. Therefore, the results provided evidence for its validity, proving to be a robust measuring instrument of teacher behavioral intentions towards the learning assessment.

Key words: Instrument construction, Instrument validation, Theory of planned behavior, Learning assessment, Classroom assessment, Teacher assessment practices.

INTRODUCCIÓN

El concepto evaluación de aprendizajes en la literatura especializada, ha ido modificándose a través del tiempo, enriqueciendo con ello su definición (Barkley y Major, 2015; Pérez Juste, 2006). De este modo, se dirige hacia una conceptualización que, junto con levantar evidencia del logro de los aprendizajes de los estudiantes, posibilita el análisis y la emisión de juicios de valor informados acerca del proceso educativo, agregando a su función certificadora original, la mejora de éste (Baas et al., 2015; Danielson, 2008, 2013; Quinquer, 2000; Sanmartí, 2007; Tejada, 1999). No obstante, los cambios en la conceptualización no necesariamente se ven reflejados en la práctica docente.

Considerando el potencial aporte de la evaluación a la calidad de los aprendizajes de los estudiantes, el Ministerio de Educación [MINEDUC] de Chile desde el año 2003 ha realizado proyectos cuyo objetivo ha sido entregar orientaciones a los docentes para efectuar una evaluación consistente con los aprendizajes promovidos por el currículo (e.g., MINEDUC, 2009a, 2009b). Sin embargo, la evidencia demuestra que la evaluación de aprendizajes ha sido persistentemente uno de los aspectos más débiles en el profesorado del país (Herrada y Zapata, 2015; MINEDUC, 2008, 2011c, 2014, 2015); situación que no sólo acontece en Chile, sino también en otras partes del mundo (Baird, 2010; DeLuca, 2012; DeLuca, Chavez, Bellara y Cao, 2013; DeLuca y Klinger, 2010; McGee y Colby, 2014; Popham, 2011). Ante la evidencia planteada, surge entonces la necesidad de contar con instrumentos de medición que permitan recabar información acerca de diferentes factores que condicionan la manera en que el profesorado efectúa el proceso evaluativo, más aun considerando su relevancia para el aprendizaje de los estudiantes.

Particularmente, este estudio se focaliza en las motivaciones intrínsecas de los docentes, medidas a través de la intención conductual: intención que éstos tienen respecto de llevar a cabo una determinada práctica evaluativa. Ajzen (1991, 2002) plantea desde la psicología social, a través de la Teoría de Conducta Planificada, que las intenciones conductuales de los sujetos se ven condicionadas por su actitud, la norma subjetiva y el control percibido con respecto a efectuar un determinado comportamiento; siendo éste el marco en el que se construye y valida el cuestionario de intenciones conductuales del profesorado hacia la evaluación de aprendizajes.

En el capítulo I se presentan los antecedentes y referentes teóricos relativos a la evaluación de aprendizajes y a la medición de variables determinantes del comportamiento volitivo, conceptos que sustentan este estudio. En el capítulo II se describe el objetivo general del presente proyecto, además de los objetivos específicos asociados a éste. Luego, en el capítulo III se detalla la metodología con la cual se construyeron, validaron y seleccionaron los ítems que conformaron el instrumento de medición en su versión piloteada y posteriormente propuesta. En el capítulo IV se da cuenta de los resultados obtenidos tras analizar psicométricamente los datos del pilotaje, siendo éstos discutidos en el capítulo V. Para finalizar el estudio, en el capítulo VI se plantean las proyecciones y limitaciones del cuestionario, respectivamente.

I. ANTECEDENTES Y REFERENTES TEÓRICOS

El presente estudio tiene su sustento teórico, primeramente, en la evaluación de aprendizajes; razón por la cual, en este capítulo se desarrolla este tópico, describiendo su evolución conceptual, características y aportes de diferentes modelos de evaluación educativa a través del tiempo, su relación con la evolución de los conceptos de aprendizaje y enseñanza, además de los antecedentes contextuales referidos a su práctica. Por otra parte, considerando que la evaluación se sitúa dentro del contexto escolar, constituyendo éste un espacio social tanto por las interacciones que allí tienen lugar como por su propósito de insertar individuos preparados en la sociedad, resulta pertinente abordar este estudio desde la psicología social; más específicamente, desde la Teoría de Conducta Planificada, desarrollándose en este capítulo la evolución de la medición de variables determinantes del comportamiento que precedieron esta teoría, junto con los aportes de ésta y su aplicación en el ámbito educativo.

1.1. Evaluación de aprendizajes

La evaluación de aprendizajes constituye un proceso en el que se recogen y analizan evidencias de aprendizaje de los estudiantes, para a partir de ello, emitir juicios de valor que orientan la toma de decisiones dentro del aula (Airasian y Russell, 2008; Jorba y Sanmartí, 2000; Sanmartí, 2007; Stiggins y Chappuis, 2012). De acuerdo a Sanmartí (2007), las decisiones derivadas de la evaluación pueden ser de carácter social o pedagógico, dependiendo de la finalidad con la que ésta se haya concebido. La evaluación tendrá una función social, si persigue constatar y certificar el nivel de logro de los alumnos, una vez finalizada una determinada etapa del proceso educativo. Por otra parte, la evaluación tendrá una función pedagógica, si tiene por objeto identificar aquellos aspectos del proceso educativo que se deben mantener, modificar o fortalecer, durante su transcurso, para conducirlo hacia su mejora.

La evaluación de aprendizajes es uno de los conceptos que más cambios ha experimentado a través del tiempo, dentro del ámbito educativo. Estas modificaciones han traído consigo un enriquecimiento conceptual, que ha llevado a la evaluación de aprendizajes desde la sola constatación de logros, hacia su uso sistemático en favor de la consecución de los objetivos de aprendizaje (Ahumada, 2005; Baas et al., 2015; Barkley y Major, 2015; Guba y Lincoln, 1989; Parra, 2008; Pérez Juste, 2006; Quinquer, 2000). Pese a ello, la evidencia empírica ha demostrado que la práctica evaluativa no ha evolucionado con igual rapidez que su conceptualización teórica (e.g., DeLuca et al., 2013; Herrada y Zapata, 2015; McGee y Colby, 2014; MINEDUC, 2014; Popham, 2011), siendo relevante el establecimiento de marcos que permitan definir concretamente los estándares de una práctica evaluativa de calidad y la identificación de los principales factores que impiden que ésta se lleve a cabo exitosamente.

1.1.1. Modelos de la evaluación de aprendizajes

Quinquer (2000) caracteriza los cambios en la conceptualización de la evaluación educativa, como una respuesta a la evolución de los conceptos de enseñanza y aprendizaje; de este modo, plantea tres modelos o enfoques: Modelo psicométrico, Modelo sistémico y Modelo comunicativo. La autora reconoce que la fundamentación teórica que sustenta esta categorización, en ocasiones, se desvincula de la evaluación que el profesorado ejerce en su sala de clases, puesto que las prácticas evaluativas “están ancladas (más a menudo de lo que sería deseable) en enfoques que tuvieron su vigencia y su razón de ser en otros momentos” (p. 14). Pese a esta disociación, resulta conveniente estudiar cada uno de estos modelos con un propósito metodológico, que permita enmarcar y comprender los cambios conceptuales de la evaluación de aprendizajes dentro de un contexto y visualizar las implicancias de ello en el quehacer educativo. En la Tabla 1, se desglosa la teoría del aprendizaje predominante en cada modelo de evaluación educativa, además de las innovaciones surgidas a partir de cada enfoque.

Tabla 1: Teorías de aprendizaje y componentes evaluativos subyacentes a los modelos de evaluación educativa

Modelos de evaluación educativa	Teorías de aprendizaje	Componentes evaluativos			
		Intencionalidad	Tipo de instrumentos	Referente	Tipo de agente
Psicométrico	Conductismo	Sumativa	Cuantitativos	Normativo Criterial	Heteroevaluación
Sistémico	Neoconductismo Cognoscitivismo	Diagnóstica* Formativa* Sumativa	Cuantitativos Cualitativos*	Normativo Criterial	Heteroevaluación
Comunicativo o psicosocial	Constructivismo	Diagnóstica Formativa Sumativa	Cuantitativos Cualitativos	Normativo Criterial Ideográfico*	Heteroevaluación Autoevaluación* Coevaluación*

Nota: *Componentes evaluativos introducidos en cada modelo de evaluación educativa.

1.1.1.1. Modelo psicométrico

Durante la primera mitad del siglo XX, el paradigma positivista y con él, los métodos de investigación basados en evidencia empírica, ejercieron una fuerte influencia en el campo de la educación; razón por la cual, en estos años surge el conductismo, primera teoría acerca del aprendizaje (Harasim, 2012). El conductismo plantea que el conocimiento sólo puede adquirirse a partir de la observación de hechos y fenómenos objetivos, poniendo especial énfasis en los estímulos externos y en las respuestas observables (Winn, 2003).

Estas corrientes de pensamiento ejercieron fuerte influencia sobre las prácticas evaluativas de la época, revelando una evaluación de aprendizajes concebida indistintamente a la medición de resultados (Ahumada, 2002). Quinquer (2000) enmarca esta concepción dentro de un enfoque de evaluación educativa que denominó el Modelo psicométrico. Bajo este modelo, la calidad de un instrumento de evaluación está dada únicamente por sus características psicométricas, sean éstas su confiabilidad y validez. En consecuencia, “las *pruebas tipo test* aparecen como la manera más «objetiva» y adecuada de medir los resultados [de] aprendizaje por su pretendida fiabilidad y por la posibilidad de cuantificar los resultados” (Quinquer, 2000, p. 15). Este enfoque asume que los resultados de aprendizaje se distribuyen de acuerdo a la curva normal, por lo cual se espera que sólo una fracción limitada del total de estudiantes alcance los puntajes más altos (Valverde, 2000). Ello posibilita la comparación y réplica de los resultados; no obstante limita el análisis a un nivel normativo (Parra, 2008).

A partir de la década de 1920, comenzaron a surgir movimientos que promovieron la formulación anticipada de objetivos de evaluación de aprendizajes (Ahumada, 2002; Parra, 2008; Quinquer, 2000). Para facilitar este proceso, se propuso la elaboración de taxonomías, las que además permitieron diversificar las fuentes de información y unificar criterios de corrección (Parra, 2008). Si bien estas innovaciones constituyeron un avance, en términos de objetividad de la evaluación y de identificación de brechas entre el logro alcanzado y el logro esperado, se mantuvo el foco en medición de resultados (Ahumada, 2002).

De acuerdo a Santos Guerra (2005), un enfoque como el antes descrito constituye una concepción simplificada de la evaluación, puesto que ignora las consecuencias asociadas a ésta, las causas del fracaso y la posibilidad de una intervención referida a procesos. Por otra parte, la concepción limitada de aprendizaje restringe los objetivos a abordar y los instrumentos de evaluación a utilizar, obviando aspectos relevantes de la formación educativa y prescindiendo de herramientas cualitativas en la recolección y análisis de datos (Ahumada, 2005; Parra, 2008; Quinquer, 2000).

Pese a ello, el Modelo psicométrico aún está presente en el contexto educativo, en tanto las prácticas de algunos docentes relevan la función social de la evaluación de aprendizajes por sobre su función pedagógica (Quinquer, 2000). El uso de pruebas objetivas como únicos instrumentos de evaluación posibles, la implementación de una evaluación centrada exclusivamente en la medición de resultados finales, y la ausencia de una retroalimentación y de una adecuación del proceso educativo en favor del aprendizaje de los estudiantes, ejemplifican la vigencia del enfoque. Ello da cuenta de la urgencia de incorporar innovaciones evaluativas a nivel de aula, que hagan un seguimiento continuo de los aprendizajes del alumnado para llevarlos a su mejora.

1.1.1.2. Modelo sistémico o comunicativo

A partir de la tercera década del siglo XX, se generaron cambios en los paradigmas científicos, los que incidieron en las concepciones de aprendizaje y ciencia; de esta manera, se plantearon nuevas orientaciones que modificaron los supuestos de base respecto de la evaluación (Quinquer, 2000). En particular, la Teoría neoconductista y la Psicología cognoscitiva, ejercieron gran influencia sobre el Modelo sistémico, introduciendo en la evaluación de aprendizajes, un carácter formativo orientado a la mejora del proceso educativo.

El Neoconductismo surge en Estados Unidos entre 1930 y 1950, manteniendo la metodología positivista del conductismo; sin embargo, legitima el planteamiento de teorías que permiten explicar y predecir comportamientos observables de los individuos (Greenwood, 2015; Mills, 2000). En materia educativa, se especifican las conductas a observar en los estudiantes una vez que el proceso educativo culmina (Sánchez y Martínez, 1998). Desde la evaluación de aprendizajes, Quinquer (2000) señala que este enfoque “da prioridad al *análisis de tareas* que han de realizar los estudiantes, se precisa su estructura, la complejidad, el grado de dificultad, así como los *prerrequisitos* o conocimientos ya adquiridos necesarios para asimilar los nuevos aprendizajes” (p. 16).

Dado el excesivo énfasis que el Neoconductismo otorgó a los estímulos externos, y a la insuficiencia con la que abordó los cambios en el paradigma científico, durante la década de 1950 surge la Psicología cognoscitiva (Seattler, 2004; Taber y Young, 2011). El Cognoscitivismo releva la importancia de los procesos internos de los estudiantes que influyen en su aprendizaje, y de cómo éstos pueden aprovecharse para promover un proceso educativo más efectivo (Mergel, 1998; Picado, 2006). Desde esta corriente, se plantean modelos que describen o explican los elementos que intervienen en el acto de aprender y la manera en que se procesa y almacena esta información (Picado, 2006).

El Neoconductismo y el Cognoscitivismo posibilitaron un cambio en la intencionalidad de las evaluaciones: “La evaluación, ahora prioritariamente *formativa* e inserta en el proceso de aprendizaje, actúa como el instrumento adecuado para *regular* y *adaptar* la programación a las *necesidades* y *dificultades* de los estudiantes” (Quinquer, 2000, p. 16). La evaluación formativa es una herramienta de intervención pedagógica al servicio de los docentes y del propio alumnado, en tanto reúne evidencia continua sobre lo que los estudiantes van aprendiendo, permitiendo adaptar tempranamente el proceso de enseñanza y aprendizaje en función de ello (Pérez Juste, 2006; Wiliam, 2011). De este modo, la evaluación de aprendizajes desde el Modelo sistémico, está presente durante todo el proceso educativo (Quinquer, 2000):

[Desde este enfoque, la evaluación permite] realizar una *diagnosis inicial* y readaptar la programación en función de los resultados, garantizar un seguimiento adecuado del proceso (*evaluación formativa*), detectando errores y dificultades, establecer diversos mecanismos de regulación (retroactiva, proactiva, interactiva), reforzar los éxitos, verificar, finalmente, los resultados mediante una *evaluación sumativa* o final (p. 16).

Habiendo reconocido en la evaluación educativa una función pedagógica orientada a la mejora del proceso educativo, ésta comienza a requerir fuentes de información que sean coherentes con este propósito; de este modo, se incorporan en el Modelo sistémico, además de los métodos cuantitativos heredados del enfoque anterior, métodos cualitativos como la observación, las entrevistas y los diarios de clase, entre otros (Pérez Juste, 2006; Quinquer, 2000).

Estableciendo semejanzas entre el Modelo sistémico y su antecesor, ambos enfoques se preocupan de la calidad de los instrumentos de evaluación empleados. Mientras que el Modelo psicométrico interpreta este requerimiento desde la medición, el Modelo sistémico lo hace desde la edumetría; es decir, analizando la calidad de las evaluaciones de acuerdo al contexto educacional, específicamente en la sala de clases (Dochy, 2009; Förster y Rojas-Barahona, 2008; Zabalza, 2007). A diferencia del supuesto de normalidad que subyace en la psicometría, desde la perspectiva edumétrica se espera que la totalidad del alumnado alcance los objetivos de aprendizaje, poniendo especial énfasis en el proceso educativo que antecede a los resultados a través del seguimiento del trabajo diario del alumnado (Quinquer, 2000; Zabalza, 2007). En este marco, Förster y Rojas-Barahona (2008) sintetizan los criterios de calidad de una evaluación, sean éstos confiabilidad y validez, redefinidos desde la edumetría. En la Tabla 2, se presenta una síntesis que da cuenta de ello, además de prácticas evaluativas sugeridas que permitirían cumplir con dichos requerimientos.

Pese a las ventajas del Modelo sistémico, éste se ha visto limitado por la complejidad de su concreción. De acuerdo a Quinquer (2000) en los centros educativos no siempre existen las condiciones laborales para que los docentes puedan planificar e implementar una evaluación formativa de manera continua y con diferentes métodos de recolección de información; especialmente considerando el número de estudiantes por sala de clases, la heterogeneidad del alumnado, o la proporción de horas lectivas de los docentes dentro de su jornada de trabajo, entre otros retos.

Tabla 2: Requerimientos técnicos de calidad desde la evaluación de aprendizajes a nivel de aula

Requerimientos	Descripción	Sugerencias en la práctica evaluativa
Validez de contenido	Cobertura y relevancia de contenidos seleccionados en un área disciplinar.	Elaborar una tabla de especificaciones del instrumento y someter la evaluación a criterio de “jueces”.
Validez instruccional	Coherencia entre lo que se va a evaluar y lo que los estudiantes han tenido oportunidad real de aprender.	Velar por que las situaciones de evaluación contengan los contenidos vistos en las actividades de aprendizaje, sean equivalentes a las actividades realizadas y utilicen un lenguaje conocido por los estudiantes.
Validez consecucional	Consecuencias que puede tener una evaluación para el aprendizaje de los estudiantes.	Definir claramente los propósitos y usos de la evaluación e identificando a priori las evidencias que darán cuenta de dichos propósitos.
Confiabilidad	Consistencia y precisión de una evaluación, vale decir, que las evidencias que se recojan del desempeño de los estudiantes sean suficientes para emitir un juicio respecto del nivel de logro de sus aprendizajes.	Evaluar en reiteradas ocasiones a un estudiante respecto de un mismo aprendizaje (instancias diferentes o en varios ítems dentro de una misma situación evaluativa); velar por la claridad de los ítems, sus instrucciones, el ambiente de aplicación y la precisión en la corrección. En cuanto a esto último, se sugiere informar previamente a los estudiantes la intencionalidad de la evaluación y los aprendizajes a evaluar, dar a conocer los criterios de evaluación, elaborar pautas de respuesta o corrección y asignar previamente los puntajes en función de la relevancia y nivel de complejidad de los aprendizajes.

Nota: Adaptado de Förster, C. y Rojas-Barahona, C. (2008). Evaluación al interior del aula: una mirada desde la validez, confiabilidad y objetividad. *Revista Pensamiento Educativo* 43(2008), 285-305.

Surge entonces la necesidad de incorporar innovaciones que respondan a estas limitaciones, de manera que sea factible llevar a cabo una evaluación de aprendizajes con una funcionalidad pedagógica; por ejemplo, responsabilizando no sólo al profesorado con tareas de evaluación, sino también desarrollando en los estudiantes competencias que permitan fomentar su autorregulación y contribuir al logro de los aprendizajes de sus pares.

1.1.1.3. Modelo comunicativo o psicosocial

A partir de las últimas décadas del siglo XX, una teoría de aprendizaje que ha desempeñado un papel importante dentro de la discusión educativa, ha sido el Constructivismo (e.g., Aznar, 1992; Carretero, 2005; Glasersfeld, 1989; Gopnik y Wellman, 2012; McPhail, 2015; Miranda, 1998; Ordoñez, 2004; Phillips, 1995; Soler, 2006). Si bien no existe un consenso con respecto a su origen, es posible identificar en sus planteamientos, la presencia de diferentes corrientes de investigación psicológica y educativa (Carretero, 2005). Impulsado por Piaget (1970) y posteriormente por Vygotsky (1978), el constructivismo considera a los estudiantes protagonistas del proceso educativo, señalando que son ellos quienes construyen y gestionan su propio conocimiento (Glasersfeld, 1989; Hoffmann, 2013; Soler, 2006). De este modo, el aprendizaje no es resultado exclusivo del ambiente ni de las disposiciones internas del alumnado, como se consideró en las teorías anteriores; sino que más bien constituye una construcción personal de quienes aprenden, “influida tanto por las *características personales* del alumno (sus esquemas de conocimiento, las ideas previas, los hábitos ya adquiridos, la motivación, las experiencias anteriores, etc.) como por el *contexto social* que se crea en el aula” (Quinquer, 2000, p. 17).

La concepción de aprendizaje del constructivismo y de otras teorías educativas contemporáneas, modificó la manera en que se percibía la pedagogía, propiciando cambios en materia de enseñanza y evaluación. Hasta finales de la década de 1970, el proceso evaluativo era considerado responsabilidad exclusiva de los docentes y, por consiguiente, externa al alumnado. En respuesta a ello, durante la década de 1980 surge el Modelo comunicativo o psicosocial, el cual incorpora a los estudiantes como agentes que regulan y evalúan sus procesos de aprendizaje, mediante la coevaluación y la autoevaluación (Quinquer, 2000). De este modo, los docentes se vuelven mediadores del proceso evaluativo, promoviendo en los estudiantes el desarrollo de habilidades reflexivas y críticas acerca de su propio aprendizaje y el de sus pares (Hoffmann, 2013; Sanmartí, 2007; Wiliam, 2011).

Al igual que su antecesor, el Modelo comunicativo considera, además de las evaluaciones con intencionalidad sumativa, evaluaciones diagnósticas y formativas, ya que existe evidencia que constata la incidencia positiva de éstas en el logro de los aprendizajes (Black et al., 2003; Jorba y Sanmartí, 2000; Quinquer, 2000). En consecuencia, este enfoque conserva una concepción de evaluación edumétrica, integrada al proceso educativo y orientada hacia la mejora de los aprendizajes. Sin embargo, la concreción del Modelo sistémico depende únicamente de las adaptaciones pedagógicas que pueda hacer el docente respecto a su enseñanza, dificultando su implementación en la praxis: el profesor por sí solo “no puede atender a la gran diversidad de obstáculos que encuentran los alumnos” (Sanmartí, 2007, p. 71). En respuesta a ello, el Modelo comunicativo traspasa responsabilidades evaluativas desde el profesor hacia sus estudiantes, aumentando notablemente las posibilidades de llevar este enfoque a la práctica “con un coste mucho menor para el docente y con más beneficio en términos de aprendizaje para los estudiantes” (Quinquer, 2000, p. 19).

De acuerdo a Wiliam (2011), la manera en que el profesor gestiona socialmente su aula, en términos de las interacciones allí acontecidas, no solo incide en la calidad de su enseñanza sino también en la calidad del proceso metacognitivo del alumno. Preguntas desafiantes, cuya respuesta no sea cerrada y que favorezcan la discusión entre estudiantes, permiten recabar información acerca del razonamiento que sostiene sus estructuras de pensamiento, sean o no erróneas, de manera de trabajar sobre ello. Al respecto, De la Torre (2002) señala que el análisis del error constituye una de las principales estrategias didácticas con la que cuenta el profesorado, pues a partir de ello los docentes obtienen información útil para guiar sus intervenciones y proveer ayuda diferenciada a sus estudiantes. Asimismo, mediante la figura de la retroalimentación, la evaluación entrega la posibilidad a docentes y pares, de orientar a quien aprende. Una retroalimentación efectiva debe estar centrada en la tarea, en cómo el estudiante la resuelve y en cómo éste puede autorregular su aprendizaje, provocando una respuesta cognitiva en el alumnado y, por consiguiente, el compromiso de la realización de acciones de mejora en el futuro (Anijovich, 2013; Wiliam, 2011).

Esta autorregulación es esencial para que los estudiantes alcancen con éxito los objetivos de aprendizaje: “Los alumnos que aprenden son fundamentalmente aquellos que saben detectar y regular ellos mismos sus dificultades y pedir y encontrar las ayudas significativas para superarlas” (Sanmartí, 2007, p. 54). De este modo, los estudiantes construyen un sistema personal de aprendizaje que se enriquece progresivamente, administrando ellos mismos sus avances y estrategias para afrontar sus tareas y los obstáculos asociados a ellas (Perrenoud, 2008; Quinquer, 2000). Esta autonomía también se enriquece con la incorporación de los pares como mediadores y agentes evaluativos: “Cuando una persona examina otros trabajos no sólo identifica las incoherencias de éstos, sino que al mismo tiempo reconoce mejor las propias” (Sanmartí, 2007, p.71).

En consecuencia, desde el Modelo comunicativo se recomienda institucionalizar en el aula estrategias orientadas a la autorregulación y a la corregulación del aprendizaje. Las actividades de clase deben ser coherentes con las situaciones evaluativas realizadas posteriormente en términos de contenido y formato, favoreciendo la apropiación gradual de la evaluación por parte del alumnado (Förster y Rojas-Barahona, 2008; Quinquer, 2000). Adicionalmente, los docentes deben promover el entendimiento común de los criterios de evaluación y el compromiso en el alumnado hacia su consecución, involucrándole en su definición (Quinquer, 2000; Sanmartí, 2007; Wiliam, 2011).

Por otra parte, con el fin de responder a la diversidad del alumnado, los referentes de contrastación desde este enfoque se amplían, concibiendo no sólo una evaluación criterial o normativa, sino también una evaluación ideográfica. En ella, los aprendizajes se evalúan en comparación a la situación de partida de cada alumno y a sus propias posibilidades de desarrollo (Casanova, 1998; Pérez Juste, 2006). Una de las ventajas de esta evaluación es que al centrarse en cada sujeto, valorando sobre todo su esfuerzo y voluntad para aprender y formarse, se promueve en el alumnado su responsabilización en el proceso de aprendizaje; sin embargo, su implementación se ve dificultada por las características del sistema educativo, donde la comparación con otros y la uniformidad de los sujetos, siguen siendo prioritarias (Casanova, 1998).

Si bien el Modelo comunicativo ha implicado un avance en materia educativa, aún es posible observar prácticas evaluativas ancladas en modelos anteriores, siendo relevante estudiar las razones que podrían explicar esta disonancia. En primer lugar, la institucionalización de un nuevo enfoque de evaluación de aprendizajes implica un cambio paradigmático que requiere tiempo (Deneen y Brown, 2016). Por ejemplo, la aceptación mayoritaria del Modelo sistémico por parte del profesorado ha significado más de 30 años, existiendo aún docentes cuya concepción de evaluación de aprendizajes está ligada a la psicometría (Ahumada, 2003). De este modo, es importante alinear la formación inicial y el perfeccionamiento docente a los planteamientos del enfoque actual, transitando desde una evaluación de aprendizajes unidireccional, que toma como único agente evaluativo al profesor, hacia una que involucra activamente al alumnado.

Por otro parte, existen factores contextuales que limitan la disposición de los docentes a implementar nuevas prácticas pedagógicas (Cornejo, 2009; Cuenca et al., 2009; Fernet, Guay, Senécal y Austin, 2012; Ruffinelli, 2014). Ejemplo de ello es la falta de materiales e instrumentos de evaluación diversificados, además de la vigencia de la certificación y de la promoción en el sistema escolar (Ahumada, 2003; Harris y Brown, 2013; Quinquer, 2000). En cuanto al contexto escolar, cabe señalar que su complejidad no siempre es abordada cabalmente por el cuerpo de conocimiento teórico. En relación a ello, Quinquer (2000) plantea una serie de obstáculos que dificultan la implementación del Modelo comunicativo en el aula, refiriéndose el primero de ellos a la excesiva amplitud de los contenidos curriculares prescritos. La saturación de contenidos no sólo implica que éstos se aborden superficialmente, limitándose a hechos memorísticos y a la reproducción de información explícita, sino que además restringe el tiempo que los docentes pueden dedicar al desarrollo de habilidades de regulación y evaluación de aprendizajes en el alumnado (Ahumada, 2003; Black, Harrison y Lee, 2004; Harris y Brown, 2013; Pérez, Soto, Sola y Serván, 2009; Santibáñez, 2002). Adicionalmente, un elevado número de estudiantes por sala de clases y por docente constituyen factores importantes a considerar, puesto que podrían dificultar el seguimiento de los educandos (Quinquer, 2000; Schwartz, Schmitt y Lose, 2012).

En síntesis, pese a que la implementación del Modelo comunicativo constituye aún un desafío para el profesorado, sus innovaciones dan cuenta de un alto valor pedagógico. Al igual que su antecesor, este enfoque caracteriza la evaluación educativa como un proceso planificado, continuo y sistemático, que cumple con criterios de rigor edumétricos, es decir, orientados a la mejora de los aprendizajes individuales de los estudiantes. Sin embargo, respondiendo a las limitaciones de los enfoques que lo precedieron, el Modelo comunicativo incorpora al alumnado como agente y mediador del proceso evaluativo, posibilitando la consecución de aprendizajes más profundos y significativos y, en consecuencia, una gestión del proceso de enseñanza más efectiva.

1.1.2. Antecedentes contextuales de las prácticas evaluativas del profesorado en la actualidad

Tal y como se desarrolló en el apartado anterior, el concepto de evaluación de aprendizajes ha experimentado diversas modificaciones a través del tiempo. No obstante, las innovaciones asociadas a su evolución conceptual no han sido incorporadas a la par en la praxis, siendo relevante estudiar el contexto en el que se enmarcan las prácticas evaluativas del profesorado en la actualidad. Particularmente, para comprender el contexto de las prácticas evaluativas en Chile, es necesario estudiar también el currículum vigente, incluyendo la concepción de evaluación que subyace en él.

De acuerdo a la Ley N° 20.370 (2009), el currículum chileno se materializa en las Bases Curriculares, conjunto de objetivos de aprendizaje que definen los desempeños mínimos esperados por los estudiantes, tanto transversalmente como por asignatura y nivel de enseñanza. Los Objetivos de Aprendizaje surgen en reemplazo de los Objetivos Fundamentales y de los Contenidos Mínimos Obligatorios, categorías establecidas en la Ley N° 18.962 (1990) en el marco de una nueva institucionalidad, la cual exigía mayor claridad y precisión en su formulación, además de una mayor alineación con la tendencia curricular internacional (MINEDUC, 2011a; 2011b).

En materia evaluativa, el currículum nacional denota una concepción de la evaluación integrada a la enseñanza, reconociendo en ella un rol central en la promoción y en el logro de los aprendizajes. Por esta razón, el MINEDUC se ha preocupado de entregar a lo largo del tiempo, herramientas para que los docentes puedan llevar a cabo el proceso evaluativo adecuadamente; ejemplo de ello es el Proyecto de Formación en Evaluación para el Aprendizaje, desarrollado en conjunto con el Programa de Educación Continua para el Magisterio [PEC] entre los años 2003 y 2009, además de las orientaciones para la evaluación, incluidas en cada uno de los documentos que conforman las Bases Curriculares vigentes (MINEDUC, 2009a, 2009b, 2012). En relación a este último instrumento, es posible identificar en él la predominancia del Modelo comunicativo de evaluación educativa (Quinquer, 2000), desarrollado en el acápite anterior, el cual adiciona a un proceso evaluativo sistemático la responsabilización del alumnado acerca de su proceso de aprendizaje.

Pese a las iniciativas antes mencionadas, la evidencia demuestra que la evaluación de aprendizajes ha sido persistentemente uno de los aspectos más débiles en el profesorado en Chile; particularmente, en los ámbitos referidos a la calidad de los instrumentos de evaluación empleados en el aula, a las interacciones profesor alumno y a la reflexión del docente acerca de la clase y de los resultados de las evaluaciones administradas (Herrada y Zapata, 2015; MINEDUC, 2008, 2011c, 2014, 2015). Esto no sólo acontece en Chile, sino también en otras partes del mundo, existiendo evidencia empírica acerca de las debilidades del profesorado, en términos de su conocimiento sobre evaluación de aprendizajes y de las dificultades que tienen para llevar a cabo prácticas evaluativas de calidad (e.g., Baird, 2010; DeLuca, 2012; DeLuca et al., 2013; DeLuca y Klinger, 2010; DeLuca, LaPointe-McEwan, Luhanga, 2015; Deneen y Brown, 2016; McGee y Colby, 2014; Popham, 2011). De esta manera, las prácticas evaluativas evidenciadas no logran alcanzar los estándares esperados para un ejercicio docente exitoso. Cabe señalar que esta situación no es generalizada, en tanto existen países como Nueva Zelanda, donde la evaluación de aprendizajes constituye una fortaleza de sus docentes (Organización para la Cooperación Económica y el Desarrollo [OECD], 2013).

En esta materia, el *Framework for Teaching Evaluation Instrument* (Danielson, 2013), constituye una herramienta reconocida a nivel mundial en el ámbito de la definición de estándares para una enseñanza de excelencia, entre los que están incluidos los referidos a las prácticas evaluativas. Este instrumento, elaborado en Estados Unidos y cuya primera versión data de 1996, se compone de cuatro dominios: Planificación y preparación, Ambiente de la clase, Enseñanza y Responsabilidades profesionales, estando presentes en todos ellos aspectos ligados a la evaluación de aprendizajes.

En el ámbito nacional, el MINEDUC en conjunto con el Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas [CPEIP] (2008), elaboraron el Marco de la Buena Enseñanza [MBE], instrumento basado en los dominios antes mencionados. Además de servir como referente en las evaluaciones de desempeño profesional docente del sector subvencionado, entre sus múltiples usos se encuentra también propiciar en el profesorado un examen de sus propias prácticas pedagógicas y la consecuente mejora de éstas.

Sin embargo, de acuerdo a un estudio realizado por la OECD (Santiago et al., 2013), la comprensión desigual del MBE por parte de los educadores, desaprovecha su potencial: las responsabilidades docentes en su conjunto, incluyendo las relacionadas a la evaluación de aprendizajes, no son concebidas de igual manera por el profesorado. Sumado a ello, debido a la ausencia de instancias formativas dentro del proceso de Evaluación del Desempeño Profesional Docente en el sector subvencionado, los profesores consideran que la retroalimentación recibida, basada en los criterios estipulados en el MBE, “no es lo suficientemente específica para otorgarle valor en relación a su práctica” (Santiago, 2013, p. 24). De este modo, aun cuando este instrumento define las prácticas docentes de calidad, incluidas las evaluativas, ello no asegura su apropiación por parte del profesorado; más aun considerando que la docencia constituye una de las profesiones más demandantes, en términos físicos, psicológicos, sociales y organizacionales (Fernet et al., 2012), especialmente en Chile (Cornejo, 2009; Cuenca et al., 2009; Ruffinelli, 2014).

En resumen, la evaluación de aprendizajes es un aspecto central dentro del currículum chileno. Si bien el MINEDUC ha implementado programas para desarrollar competencias evaluativas en el profesorado, este tópico ha demostrado ser una de las principales debilidades del desempeño docente en Chile, repitiéndose esta situación en otras partes del mundo. El conocimiento evidenciado por los docentes sobre la evaluación de aprendizajes es insuficiente, al igual que la comprensión sobre el desempeño que se espera de ellos en este ámbito, lo cual se reflejaría en la calidad de sus instrumentos evaluativos, en las interacciones acontecidas en el aula y en sus procesos reflexivos. Con respecto a ello, Deneen y Brown (2016) plantean que la alfabetización en materia evaluativa y las concepciones afectivas de los profesores sobre la evaluación de aprendizajes inciden en la manera en que este proceso se lleva a la práctica. En consecuencia, resulta fundamental estudiar más profundamente la relación entre el profesorado y la evaluación educativa.

1.2. Medición de variables determinantes del comportamiento

Las prácticas evaluativas del profesorado están sujetas a diversos factores contextuales, los que condicionan su calidad y el modelo evaluativo en el cual se suscriben. Sin embargo, no sólo variables externas a los docentes influyen en su praxis. Reconociendo la incidencia de un factor intrínseco, De la Torre (2002) plantea que el contexto socioeducativo en el que se desempeña el docente se ve afecto por un filtro personal, siendo relevante distinguir los componentes internos que facilitan u obstaculizan la implementación de nuevos enfoques evaluativos en el aula: “Como el bagaje cultural de un pueblo al ser asimilado por los miembros de la comunidad, cada uno integra la cultura objetiva de forma personal; personaliza y recrea la cultura. Cada profesor, integra a su modo esa cultura sobre la evaluación” (p. 263). Ante este escenario, surge la necesidad de contar con instrumentos de medición que permitan recabar información acerca del conjunto de componentes intrínsecos a los docentes que los motivan a llevar a cabo sus prácticas evaluativas de determinado modo.

Considerando que la evaluación de aprendizajes se encuentra inserta dentro del contexto escolar, constituyendo éste un espacio social, tanto por las interacciones que allí tienen lugar como por su propósito de preparar individuos para la sociedad, resulta pertinente abordar este estudio desde la psicología social. Esta disciplina estudia la influencia que tiene sobre el pensamiento, sentimiento y comportamiento de los individuos, la presencia real, imaginada o implícita de otras personas (Allport, 1985; Fiske, 2010; Franzoi, 2010).

1.2.1. Medición de actitudes

Un concepto que ha desempeñado un papel importante en el desarrollo de la psicología social, es el de actitud, definida como “tendencia psicológica expresada mediante la evaluación de una entidad particular con cierto grado de aprobación o desaprobación” (Eagly y Chaiken, 1993, p. 1). Este término comenzó a emplearse a fines del siglo XIX en la literatura especializada y se le atribuyó una incidencia directa sobre el comportamiento de las personas, afirmación basada en escaso fundamento empírico (Ajzen y Fishbein, 1980).

Si bien en la primera mitad del siglo XX, los estudios de medición de actitudes caracterizaron este constructo de manera unidimensional, existían definiciones contrapuestas atribuidas al mismo concepto, dando cuenta de la necesidad de conceptualizarlo más rigurosamente y de capturar así su complejidad (Ajzen y Fishbein, 1980; 2005). En consecuencia, Allport (1935) propuso medir y analizar la actitud, desagregando el constructo en unidades más pequeñas pero con suficiente significado. En respuesta a esta recomendación, algunos autores (e.g. Cartwright, 1949; Katz y Stotland, 1959; Smith, 1947) comenzaron a incorporar en sus estudios una concepción multidimensional de actitud, resultando común hallar entre sus componentes, dimensiones cognitivas, afectivas y conductuales (Ajzen y Fishbein, 1980; Eagly y Chaiken, 1993).

1.2.1.1. Modelo de tres componentes de actitud

En 1960, Rosenberg y Hovland elaboraron una representación esquemática de la configuración múltiple de actitud (Fig. 1). En este modelo, actitud se define como la predisposición a responder de una manera determinada a un estímulo específico, pudiendo agruparse estas respuestas en tres componentes: cognitivas, afectivas y conductuales. El componente cognitivo da cuenta de los pensamientos que las personas tienen acerca del objeto evaluado; el componente afectivo, de los sentimientos y emociones que presentan hacia éste; y el componente conductual, de las acciones que tienden a llevar a cabo con respecto a él (Eagly y Chaiken, 1993).

Aun cuando este modelo ha sido empleado ampliamente en diversas investigaciones (e.g., Brahm y Euler, 2013; Kim, Lu y Estrada-Hernandez, 2015; Zabalza, 1998), la incidencia de la actitud definida bajo esta concepción se sustenta en escasa evidencia empírica (Ajzen y Fishbein, 1980). Ello revela la necesidad de plantear nuevos enfoques; como por ejemplo, el propuesto por Bagozzi y Burnkrant (1979), en el que las dimensiones afectiva y cognitiva podrían determinar la predisposición conductual de los sujetos.

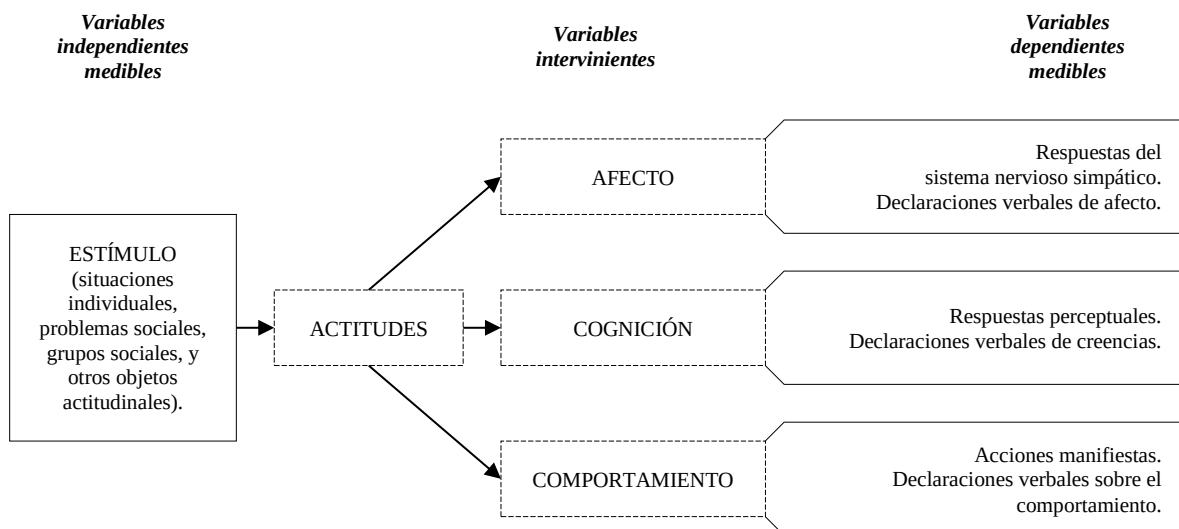


Fig. 1: Modelo de tres componentes de actitud

1.2.2. Medición de la intención conductual

Ante la necesidad de plantear modelos que incorporen además de la actitud otros factores que intervienen en la decisión de actuar de una manera en particular, surge la Teoría de Acción Razonada (Ajzen y Fishbein, 1980) y, posteriormente, la Teoría de Conducta Planificada (Ajzen, 1991, 2002). Los autores de ambas teorías afirman que gran parte de los comportamientos sociales están sujetos al control volitivo y, por tanto, pueden predecirse a partir de las intenciones. En este marco, la intención conductual es una variable que captura, dentro de un plan consciente, el esfuerzo que una persona decide que pondrá en realizar un determinado comportamiento; no obstante, se distinguen entre sí por los componentes (Eagly y Chaiken, 1993).

1.2.2.1. Teoría de la Acción Razonada

La Teoría de la Acción Razonada (Fig. 2) constituye un modelo que vincula las creencias de los individuos con sus actitudes, la manera en que perciben las expectativas del resto y sus intenciones, estando estas últimas relacionadas estrechamente con la toma de decisiones a nivel conductual (Ajzen y Fishbein, 1980). En esta teoría, la intención se compone no sólo de las actitudes de los sujetos hacia la realización de la acción evaluada, sino también de la norma subjetiva; “esto es, los individuos intentan realizar una conducta cuando evalúan ésta como positiva y creen que otros, el otro importante generalizado, es decir las personas y los grupos socialmente relevantes para el sujeto, consideran que ellos deben realizarla” (Reyes, 2007, p. 70).

Los instrumentos enmarcados en esta teoría abordan la intención conductual tanto desde sus componentes, la actitud y la norma subjetiva, como desde la propia intención conductual, considerada directamente mediante la disposición de los individuos a efectuar una determinada acción (Eagly y Chaiken, 1993; Reyes, 2007). De esta manera, se facilita el estudio sobre la relación de cada componente con la intención, y a su vez la relación de esta última con la conducta.

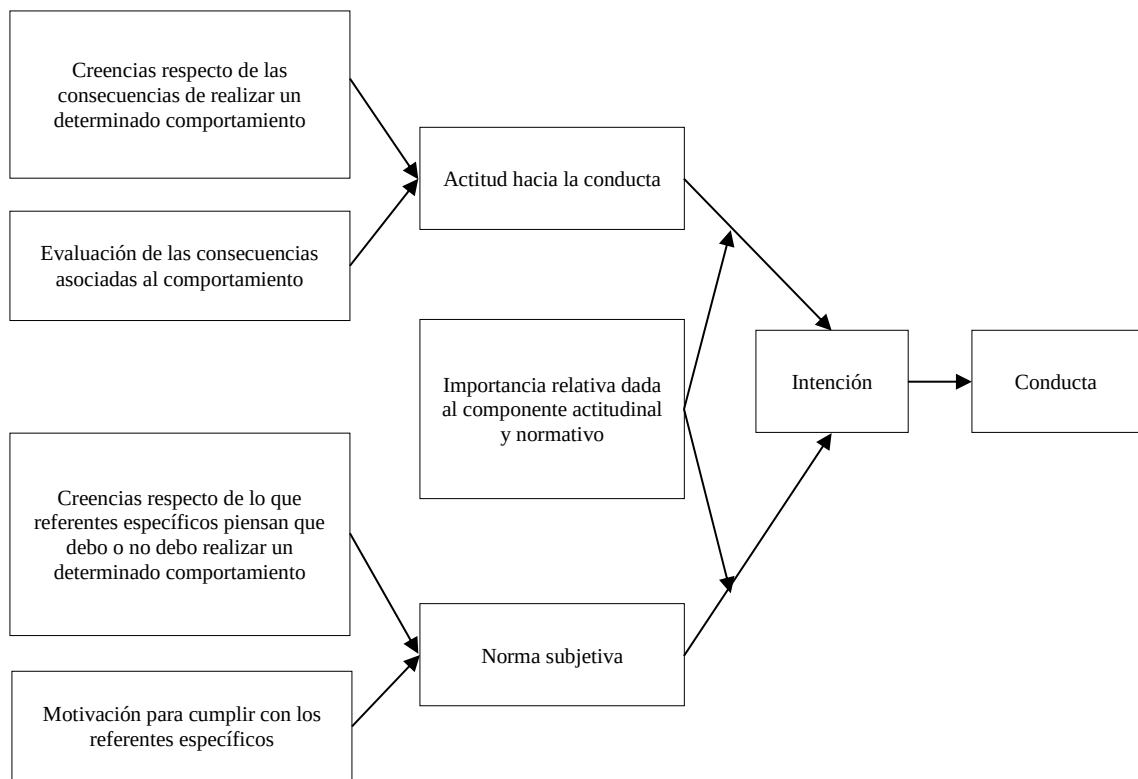


Fig. 2: Teoría de la Acción Razonada

Pese a que la Teoría de la Acción Razonada cuenta con robusta evidencia empírica que avala su aplicación en diversos contextos, ha recibido críticas respecto a la excesiva simplificación de las variables involucradas en la toma de decisiones; situación que la Teoría de Conducta Planificada busca revertir mediante la incorporación de un tercer componente en el modelo (Eagly y Chaiken, 1993).

1.2.2.2. Teoría de Conducta Planificada

Considerando las críticas del modelo anterior, Ajzen (1991, 2002) plantea la Teoría de Conducta Planificada (Fig. 3), la cual agrega a la estructura de su antecesor, un tercer componente de las intenciones conductuales: el control percibido, definido como la percepción acerca de cuán fácil o difícil es llevar a cabo una acción.

Al igual que en el modelo anterior, los instrumentos enmarcados en esta teoría abordan la intención conductual directa e indirectamente; es decir, no solo estudian la intención conductual desde la disposición a efectuar una conducta determinada, sino que adicionalmente recaban información acerca de este constructo desde sus componentes, sean éstos actitud, norma subjetiva y control conductual percibido (Ajzen, 2002; Eagly y Chaiken, 1993; Yan y Cheng, 2015). La distinción entre estos elementos permite examinar la relación entre cada uno de ellos, además del potencial efecto que podrían tener sobre las conductas finalmente realizadas por los individuos.

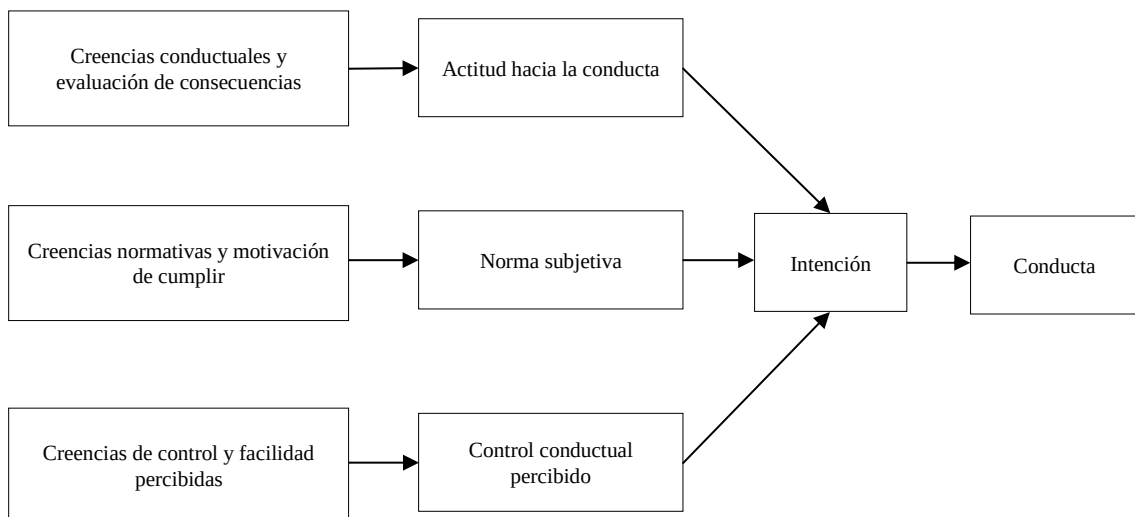


Fig. 3: Teoría de Conducta Planificada

La Teoría de la Conducta Planificada ha sido empleada en diversas áreas, existiendo más de 1.200 investigaciones académicas en revistas de punta que aluden a ella (McCoy, 2014). En educación, esta teoría se ha aplicado en el ámbito de la incorporación de tecnología en el aula (e.g., Lee, Cerreto y Lee, 2010; Teo y Lee, 2010), la inclusión de estudiantes con discapacidad (e.g., Conatser, Block y Gansneder, 2002; Stanovich y Jordan, 1998; Theodorakis, Bagiatis y Goudas, 1995), la implementación de nuevas metodología de enseñanza (e.g., Beck, Czerniak y Lumpe, 2000; Crawley, 1990; Martin y Kulinna, 2004) y la realización de determinadas prácticas evaluativas. A continuación se presentan ejemplos de este último tópico.

El estudio realizado por Yan y Cheng (2015) explora la relación entre actitudes, intenciones y prácticas docentes referidas a la evaluación formativa, en una muestra de 450 profesores de 10 escuelas primarias de Hong Kong. Este instrumento fue elaborado a partir de una revisión bibliográfica acerca de la Teoría de Conducta Planificada, además de consultas con expertos en evaluación educativa y profesores en ejercicio. De acuerdo a los resultados del estudio, aun cuando los componentes de las intenciones resultaron ser óptimos predictores de ésta, las intenciones no predijeron las prácticas evaluativas que los profesores declaran llevar a cabo. Los autores atribuyeron este hecho a la presencia de factores externos que obstaculizan la concreción de las prácticas evaluativas formativas intencionadas, como las características del grupo curso o un sistema escolar altamente vinculado a la rendición de cuentas. Una de las limitaciones de este estudio radica en que sólo se consideran prácticas referidas a las evaluaciones de intencionalidad formativa, habiendo acciones que enmarcan la evaluación de aprendizajes y que no solo corresponden a dicho propósito, como lo son la planificación de la evaluación, la comunicación de resultados, la toma de decisiones de carácter sumativo, entre otras.

Un estudio que sí considera la evaluación en su conjunto, es el realizado por Brown et al. (2009), el cual toma como hipótesis la premisa que fundamenta la Teoría de Conducta Planificada, en tanto un constructo intrínseco del profesorado determina su praxis. Para ello, estos autores examinaron la relación existente entre las concepciones de cerca de 300 docentes de Hong Kong, respecto de la evaluación de aprendizajes y sus prácticas evaluativas. En primer lugar, se tradujo y adaptó desde el inglés al chino, el instrumento *Teachers' Conceptions of Assessment abridged Inventory [TCoA-III]*. Este instrumento fue elaborado por Brown (2006) y sus dimensiones corresponden a la evaluación a grupos de concepciones evaluativas, sean éstas la mejora del proceso educativo, la rendición de cuentas de docentes y alumnos, y la irrelevancia del proceso evaluativo¹. En segundo lugar, se construyó y validó el instrumento *Practices of Assessment Inventory [PrAI]*, cuyas dimensiones concordaban con las del instrumento anterior. Los resultados obtenidos en el estudio indicaron que las creencias de los profesores sobre la naturaleza y el propósito de la evaluación de aprendizajes, estaban fuertemente alineadas a las prácticas evaluativas que declaran realizar. Si bien los autores del estudio señalan que su marco conceptual está basado en la Teoría de Conducta Planificada, no se explicita la relación entre los reactivos que forman parte de los instrumentos empleados y los componentes que conforman la teoría; por tanto, el estudio no presentaría una aplicación directa de la misma, dando espacio a la construcción y validación de un nuevo instrumento que recoja de manera más precisa la información acerca del constructo de interés.

¹ El instrumento *TCoA-III*A originalmente fue elaborado en inglés para medir las concepciones de los profesores de primaria y secundaria neozelandeses (Brown, 2004; Brown, 2006; Brown, 2011; Brown, Lake y Matters, 2011). Ha sido traducido, adaptado y validado en diferentes idiomas; e.g., chino en Hong Kong (Brown et al., 2009), turco en Turquía (Vardar, 2010), griego en Chipre (Brown y Michaelides, 2011), español en España (Brown y Remesal, 2012), árabe en Egipto (Gebril y Brown, 2014) y recientemente en inglés con asistencia en hindi en India (Brown, Chaudhry y Dhamija, 2015).

II. OBJETIVOS

2.1. Objetivo general

Construir y validar un instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes, dirigido a profesores de Enseñanza Básica y Media del sistema escolar chileno, que estén ejerciendo docencia en aula.

2.2. Objetivos específicos

- Construir una versión preliminar del instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes.
- Analizar la evidencia a favor de la validez del instrumento, mediante un análisis factorial de las respuestas obtenidas en su pilotaje.
- Determinar si los ítems que constituyen el instrumento, presentan diferencias significativas entre grupos de acuerdo a variables género y región.
- Determinar la capacidad discriminativa de los ítems que constituyen el instrumento.
- Determinar la consistencia interna del instrumento y de sus dimensiones.
- Proponer una versión validada del instrumento, basada en la información psicométrica obtenida en el pilotaje de su versión preliminar.

III. METODOLOGÍA

3.1. Construcción de ítems

De acuerdo a los objetivos de este estudio, el primer paso a efectuar consistió en la construcción de los ítems que conformaron el instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes.

En primer lugar, se llevó a cabo una revisión bibliográfica acerca de la evaluación de aprendizajes, con objeto de identificar prácticas asociadas a procesos evaluativos de calidad; particularmente, enmarcados en los planteamientos del Modelo comunicativo de evaluación de aprendizajes planteado por Quinquer (2000). A partir de ello, se definieron cinco categorías de prácticas evaluativas, detalladas en la Tabla 3: Obtención de información edumétricamente válida y confiable, Incorporación activa del alumnado en los procesos evaluativos, Ejecución de un proceso evaluativo planificado, continuo y sistemático, Retroalimentación orientada hacia la mejora en los aprendizajes y Responsabilización docente dentro del proceso evaluativo. Paralelamente, se realizó una revisión bibliográfica acerca de la Teoría de Conducta Planificada (Ajzen, 1991, 2002), lo cual derivó en la definición de tres escalas presentadas en detalle en la Tabla 4.

Considerando lo anterior, se formularon 684 reactivos de selección de respuesta formato Likert, con una escala de 4 puntos que va de “Muy de acuerdo” a “Muy en desacuerdo”; en consecuencia, el nivel de medición del instrumento es ordinal (Hogan, 2004; Ritchey, 2002). Para la elaboración de estos ítems, se consideraron directrices sugeridas por Haladyna (1999), Hogan (2004) y Likert (1932), resguardando que se ajusten a las definiciones operacionales de cada escala, evitando inducir respuestas socialmente deseables y procurando mantener la sencillez y claridad en la redacción de las afirmaciones. Se decidió desarrollar algunos ítems redactados inversamente, es decir, cuyo grado de acuerdo denote ausencia del rasgo medido, pues aún no existe consenso entre quienes estudian la formulación de este tipo de ítems (e.g., DiStefano y Motl, 2006; Lindwall et al., 2012), respecto de si constituyen o no sesgo de método.

Tabla 3: Categorías consideradas en la construcción de ítems de acuerdo al Modelo comunicativo de evaluación de aprendizajes

Categorías	Definición
Obtención de información edumétricamente válida y confiable	Prácticas docentes referidas a la obtención de evidencias de aprendizaje válidas y confiables, en términos de la cobertura y relevancia de los contenidos evaluados, de la coherencia entre lo que se evalúa y lo que se enseña, de las consecuencias asociadas a la evaluación y de la consistencia y precisión de una evaluación.
Incorporación activa del alumnado en los procesos evaluativos	Prácticas docentes referidas a la incorporación activa del alumnado en el proceso evaluativo, como agentes que regulan y evalúan su aprendizaje y el de sus pares, mediante el desarrollo de habilidades reflexivas y críticas acerca de su propio aprendizaje y el de sus pares.
Ejecución de un proceso evaluativo planificado, continuo y sistemático	Prácticas docentes referidas a la ejecución de una evaluación de aprendizajes planificada, continua y sistemática, inserta en el proceso educativo, en diferentes momentos y con diferentes intencionalidades, permitiendo a través de ello la adaptación de la enseñanza a las necesidades de los estudiantes.
Retroalimentación orientada hacia la mejora en los aprendizajes	Prácticas docentes referidas a una retroalimentación que orienta a quien aprende, que se centra en la tarea, en cómo el estudiante la resuelve y en cómo éste puede autorregular su aprendizaje, provocando una respuesta cognitiva en el alumnado y, por consiguiente, el compromiso de la realización de acciones de mejora en el futuro.
Responsabilización docente dentro del proceso evaluativo	Prácticas docentes referidas a la reflexión sistemática sobre su quehacer, asumiendo responsabilidades en la orientación de sus alumnos, fomentando sus fortalezas y promoviendo la superación de sus debilidades, en términos de la consecución de los aprendizajes propuestos.

La Tabla 5 muestra la distribución de ítems, considerando el cruce entre la información referida al tipo de práctica evaluativa abordada en el ítem y a la variable psicológica implicada en su formulación.

Tabla 4: Escalas y subescalas consideradas en la construcción de ítems de acuerdo a la Teoría de Conducta Planificada

Escala	Subescala	Definición
Componentes de la Intención conductual	Actitud cognitiva	Tendencia psicológica expresada mediante la evaluación de una entidad particular con cierto grado de aprobación o desaprobación. Es actitud cognitiva, en tanto constituyen los pensamientos que las personas tienen acerca del objeto evaluado.
	Actitud afectiva	Tendencia psicológica expresada mediante la evaluación de una entidad particular con cierto grado de aprobación o desaprobación. Es actitud afectiva, en tanto contempla los sentimientos y emociones que las personas presentan hacia el objeto evaluado.
	Norma subjetiva desde los superiores	Razonamiento normativo que refleja la influencia que tiene el entorno social sobre el sujeto en cuestión, respecto a si éste debe o no realizar determinada conducta. En este caso, la norma está dictada por los superiores del docente dentro de la institución educativa en la que éste se desempeña.
	Norma subjetiva desde pares	Razonamiento normativo que refleja la influencia que tiene el entorno social sobre el sujeto en cuestión, respecto a si éste debe o no realizar determinada conducta. En este caso, la norma está dictada por los pares del docente dentro de la institución educativa en la que éste se desempeña; es decir, otros docentes del establecimiento.
	Control percibido limitado por las competencias	Percepción del individuo acerca cuán capaz es de realizar determinada conducta. En este caso, su capacidad se encuentra limitada por las competencias evaluativas que posee el docente.
	Control percibido limitado por tiempo	Percepción del individuo acerca cuán capaz es de realizar determinada conducta. En este caso, su capacidad se encuentra limitada por el tiempo que dispone el docente.
Intención conductual		Disposición del individuo a realizar a futuro determinada conducta.
Conducta		Respuesta declarada por el individuo, respecto a la realización de determinada conducta en el último tiempo.

Tabla 5: Distribución de ítems contruidos de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada

Categorías de prácticas evaluativas del Modelo comunicativo							
Escalas de la Teoría de Conducta Planificada	Subescalas de la Teoría de Conducta Planificada	Obtención de información edumétricamente válida y confiable	Incorporación activa del alumnado en los procesos evaluativos	Ejecución de un proceso evaluativo planificado, continuo y sistemático	Retroalimentación orientada hacia la mejora en los aprendizajes	Responsabilización docente dentro del proceso evaluativo	Total
Componentes de la Intención conductual	Actitud cognitiva	34 ítems	34 ítems	20 ítems	25 ítems	30 ítems	143 ítems
	Actitud afectiva	33 ítems	19 ítems	17 ítems	16 ítems	14 ítems	99 ítems
	Norma subjetiva desde los superiores	23 ítems	15 ítems	14 ítems	11 ítems	12 ítems	75 ítems
	Norma subjetiva desde pares	24 ítems	13 ítems	16 ítems	11 ítems	12 ítems	76 ítems
	Control percibido limitado por las competencias	24 ítems	11 ítems	16 ítems	10 ítems	12 ítems	73 ítems
	Control percibido limitado por tiempo	21 ítems	11 ítems	16 ítems	10 ítems	12 ítems	70 ítems
Intención conductual		19 ítems	11 ítems	15 ítems	17 ítems	12 ítems	74 ítems
Conducta		19 ítems	11 ítems	15 ítems	17 ítems	12 ítems	74 ítems
Total		197 ítems	125 ítems	129 ítems	117 ítems	116 ítems	684 ítems

3.2. Validación de ítems

Una vez contruidos los reactivos, se procedió a su validación. La primera fase de esta validación, permitió seleccionar los ítems que conformaron la versión piloto del instrumento; mientras que la segunda, consistió en la evaluación de sus propiedades psicométricas a raíz de esta aplicación masiva. Estas etapas se describen a continuación.

3.2.1. Prepiloteaje

3.2.1.1. Juicio de expertos

En primer lugar, se procedió a validar el contenido de las afirmaciones contruidas mediante el juicio de cinco expertos: dos expertos en evaluación de aprendizajes, dos en psicología social y uno en construcción de instrumentos de medición. Cada uno recibió digitalmente los ítems elaborados, además de las definiciones operacionales de las escalas que guiaron la construcción. Con dicha información, los jueces señalaron si aceptaban, rechazaban o modificaban cada ítem. De los 684 reactivos contruidos, 387 presentaron una aprobación igual o superior al 80%, cumpliéndose con ello el criterio de Polit, Beck y Owen (2007) (ver Anexo 1). La Tabla 6 muestra la distribución de estos ítems de acuerdo a las prácticas evaluativas subyacentes y a las variables psicológicas de la Teoría de Conducta Planificada.

Posteriormente, se revisaron cualitativamente los comentarios de expertos, considerando posibles modificaciones de los ítems en función de su correspondencia con la definición operacional de las escalas, la claridad de su redacción y la minimización de la deseabilidad social de las respuestas. Adicionalmente, tras una sugerencia general de los expertos en evaluación de aprendizajes, se fusionaron las categorías Ejecución de un proceso evaluativo planificado, continuo y sistemático y Responsabilización docente dentro del proceso evaluativo, dado que la primera categoría forma parte de la segunda, constituyendo una responsabilidad del profesorado.

Tabla 6: Distribución de ítems con porcentaje de aprobación en juicio de expertos igual o superior a 80%, de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada

Categorías de prácticas evaluativas del Modelo comunicativo							
Escalas de la Teoría de Conducta Planificada	Subescalas de la Teoría de Conducta Planificada	Obtención de información edumétricamente válida y confiable	Incorporación activa del alumnado en los procesos evaluativos	Ejecución de un proceso evaluativo planificado, continuo y sistemático	Retroalimentación orientada hacia la mejora en los aprendizajes	Responsabilización docente dentro del proceso evaluativo	Total
Componentes de la Intención conductual	Actitud cognitiva	15 ítems	9 ítems	8 ítems	12 ítems	15 ítems	59 ítems
	Actitud afectiva	10 ítems	3 ítems	4 ítems	6 ítems	5 ítems	28 ítems
	Norma subjetiva desde los superiores	12 ítems	6 ítems	13 ítems	4 ítems	9 ítems	44 ítems
	Norma subjetiva desde pares	11 ítems	10 ítems	10 ítems	6 ítems	10 ítems	47 ítems
	Control percibido limitado por las competencias	19 ítems	7 ítems	11 ítems	6 ítems	9 ítems	52 ítems
	Control percibido limitado por tiempo	16 ítems	11 ítems	14 ítems	5 ítems	8 ítems	54 ítems
Intención conductual		8 ítems	5 ítems	8 ítems	5 ítems	7 ítems	33 ítems
Conducta		19 ítems	9 ítems	14 ítems	17 ítems	11 ítems	70 ítems
Total		110 ítems	60 ítems	82 ítems	61 ítems	74 ítems	387 ítems

3.2.1.2. Pretest cognitivo

En base a lo anterior, y resguardando la cobertura homogénea tanto de las escalas psicológicas como de las categorías asociadas a las prácticas evaluativas, se seleccionaron 100 ítems, cuya distribución se presenta en la Tabla 7. Para llevar a cabo esta selección, se optó por aquellos reactivos que mejor representaban los constructos estudiados, velando que las afirmaciones elegidas no fuesen redundantes entre sí.

Para obtener mayor evidencia a favor de la validez semántica y de contenido de los reactivos, se efectuó un pretest cognitivo, utilizando la técnica *respondent debriefing* (Hess y Singer, 1995; Hughes, 2004). Por ello, se administró un cuestionario conformado por los ítems seleccionados a una muestra de ocho profesores de Enseñanza Básica y Media, de diferentes disciplinas, con características similares a las de la población a la que dirigía el instrumento. Simulando las condiciones de la aplicación masiva, el instrumento fue respondido virtualmente; además, se les dio a conocer a los participantes el objetivo de esta aplicación mediante un consentimiento informado (ver Anexo 2). Tras responder el cuestionario, cada uno de los encuestados dio cuenta de su experiencia con el instrumento a través de una entrevista, permitiendo recabar información acerca de la manera en que ellos interpretaban los ítems respondidos y, a partir de ello, modificar la redacción de algunos reactivos o la estructura original del cuestionario. El tiempo de respuesta promedio de los encuestados fue de 26 minutos, con un mínimo de 20 minutos y un máximo de 33 minutos.

Considerando los comentarios de los respondentes, se decidió reducir la cantidad de ítems del cuestionario de 100 a 80 ítems, incorporando modificaciones en la redacción de los reactivos y prescindiendo de aquéllos que presentasen dificultades, ya sea por falta de claridad o de pertinencia como por reiterar prácticas evaluativas mencionadas en otros ítems (ver Anexo 3). La Tabla 8 presenta la distribución de estos ítems, nuevamente de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a las variables psicológicas de la Teoría de Conducta Planificada.

Tabla 7: Distribución de ítems seleccionados para pretest cognitivo, de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada

Escala de la Teoría de Conducta Planificada	Subescalas de la Teoría de Conducta Planificada	Categorías de prácticas evaluativas del Modelo comunicativo				Total
		Obtención de información edumétricamente válida y confiable	Incorporación activa del alumnado en los procesos evaluativos	Retroalimentación orientada hacia la mejora en los aprendizajes	Responsabilización docente dentro del proceso evaluativo	
Componentes de la Intención conductual	Actitud cognitiva	2 ítems	2 ítems	3 ítems	3 ítems	10 ítems
	Actitud afectiva	3 ítems	3 ítems	2 ítems	2 ítems	10 ítems
	Norma subjetiva desde los superiores	3 ítems	2 ítems	3 ítems	2 ítems	10 ítems
	Norma subjetiva desde pares	2 ítems	3 ítems	2 ítems	3 ítems	10 ítems
	Control percibido limitado por las competencias	2 ítems	3 ítems	3 ítems	2 ítems	10 ítems
	Control percibido limitado por tiempo	3 ítems	2 ítems	2 ítems	3 ítems	10 ítems
Intención conductual		5 ítems	5 ítems	5 ítems	5 ítems	20 ítems
Conducta		5 ítems	5 ítems	5 ítems	5 ítems	20 ítems
Total		25 ítems	25 ítems	25 ítems	25 ítems	100 ítems

Tabla 8: Distribución de ítems seleccionados tras realización de pretest cognitivo, de acuerdo al conjunto de prácticas evaluativas del Modelo comunicativo y a variables psicológicas de la Teoría de Conducta Planificada

Escala de la Teoría de Conducta Planificada	Subescalas de la Teoría de Conducta Planificada	Categorías de prácticas evaluativas del Modelo comunicativo				Total
		Obtención de información edumétricamente válida y confiable	Incorporación activa del alumnado en los procesos evaluativos	Retroalimentación orientada hacia la mejora en los aprendizajes	Responsabilización docente dentro del proceso evaluativo	
Componentes de la Intención conductual	Actitud cognitiva	2 ítems	2 ítems	2 ítems	2 ítems	8 ítems
	Actitud afectiva	2 ítems	2 ítems	2 ítems	2 ítems	8 ítems
	Norma subjetiva desde los superiores	2 ítems	2 ítems	2 ítems	2 ítems	8 ítems
	Norma subjetiva desde pares	2 ítems	2 ítems	2 ítems	2 ítems	8 ítems
	Control percibido limitado por las competencias	2 ítems	2 ítems	2 ítems	2 ítems	8 ítems
	Control percibido limitado por tiempo	2 ítems	2 ítems	2 ítems	2 ítems	8 ítems
Intención conductual		4 ítems	4 ítems	4 ítems	4 ítems	16 ítems
Conducta		4 ítems	4 ítems	4 ítems	4 ítems	16 ítems
Total		20 ítems	20 ítems	20 ítems	20 ítems	80 ítems

3.2.2. Pilotaje

3.2.2.1. Participantes

La población destinataria del instrumento de medición, corresponde a profesores de Enseñanza Básica y Media del sistema escolar chileno, que durante el 2015 hayan estado ejerciendo docencia en aula.

La muestra fue seleccionada mediante muestreo no probabilístico; particularmente, muestreo mixto, que combinó la autoselección de sus participantes con un muestreo en cadena (Hernández, Fernández y Baptista, 2010). Se invitó a participar de este estudio mediante correo electrónico a egresados de Pedagogía Básica y Media y a estudiantes de Postgrado en Educación, tanto de la Pontificia Universidad Católica de Chile como de la Universidad de Chile, que estuvieran ejerciendo docencia en aula. También se invitó a participar mediante redes sociales a profesores que cumplieren con dicha condición, de manera de tener cobertura en todo el país, indistintamente la universidad de formación de Pregrado o Postgrado. Adicionalmente, se instó a los participantes, a difundir entre sus colegas en ejercicio².

La muestra a la que se aplicó el instrumento de medición se compuso de 509 participantes, cifra que representa cerca del 0,23% de la población de docentes en Chile que al 2015 se desempeñaba en aula (Elige Educar, 2015). Pese a no ser una muestra aleatoria, se verificó mediante un Test Chi-cuadrado la representatividad de la muestra respecto de la población, en términos de género de los docentes, años de experiencia y dependencia del establecimiento en el que se desempeñaban³. De este modo, no se rechazó la hipótesis de que los datos se ajustasen a la distribución poblacional en las variables antes mencionadas (ver Anexo 4).

² Se informó a todos los participantes el objetivo de este estudio, siendo su participación voluntaria. Como incentivo, se sortearon cuatro *giftcards*: dos para los que respondieron y dos para los que difundieron el cuestionario. Sólo un tercio de los participantes accedió a concursar de este sorteo.

³ Para comparar los datos de la población con los de la muestra, se recodificó la variable *años de experiencia*, en función de la información publicada por Elige Educar (2015).

La muestra se conformó por 401 mujeres (78,8%) y 108 hombres (21,2%). En cuanto a su distribución geográfica, el 63,1% de ellos trabajó durante el 2015 en establecimientos de la Región Metropolitana, correspondiendo el 36,9% restante a docentes que trabajaron en otras regiones del país. En relación a la dependencia, la mayor parte de los encuestados se desempeñó durante el 2015 en el sector particular subvencionado (46,0%) o en el sector municipal (36,1%). Del total de encuestados, el 46,8% se desempeñó únicamente en Enseñanza Básica; mientras que el 19,6%, únicamente en Enseñanza Media. La Fig. 4 da cuenta de la distribución porcentual de los participantes, respecto de la asignatura impartida.

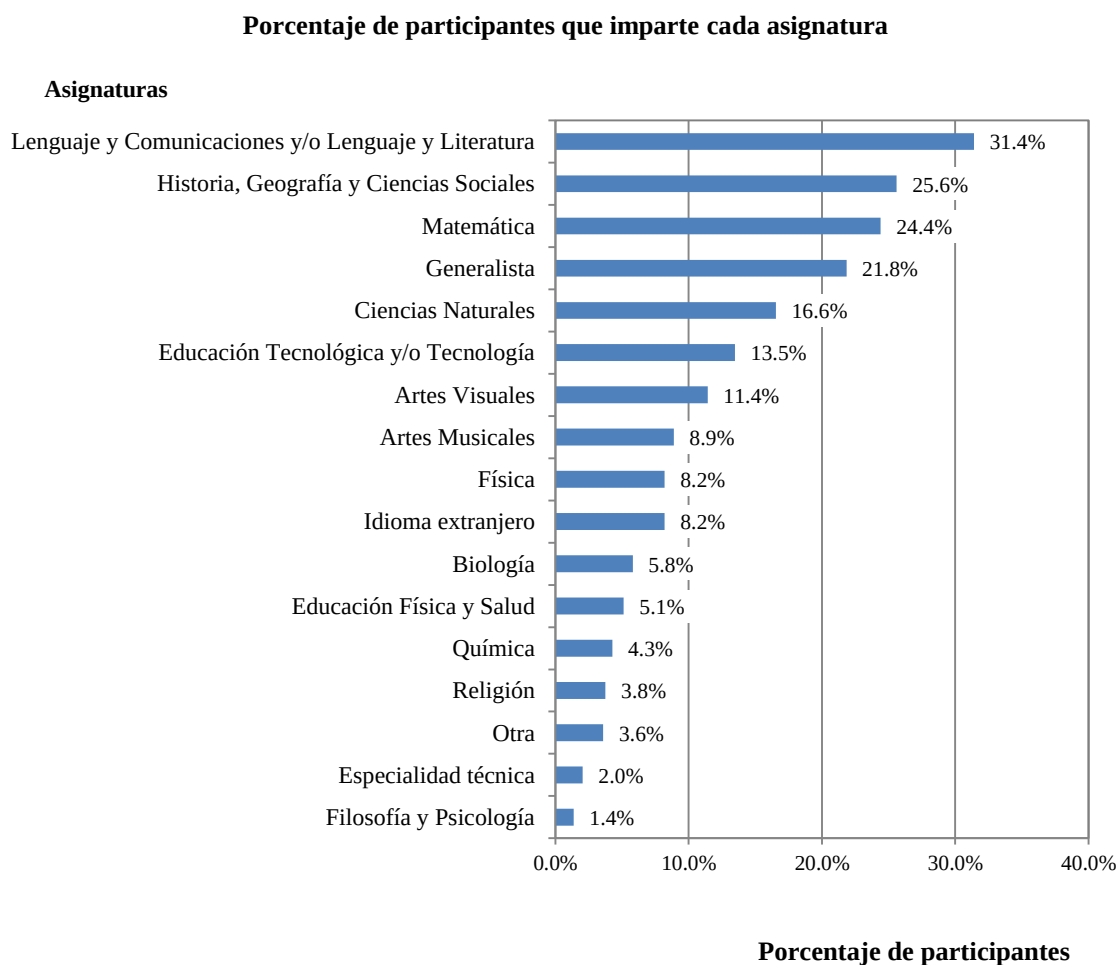


Fig. 4: Porcentaje de participantes que imparte cada asignatura

El promedio de horas lectivas de los encuestados fue de 30 horas (*d. s.* = 9). Por otra parte, el 60,5% del total de ellos presentó una razón entre horas lectivas y no lectivas superior a 80:20. En promedio, los docentes declararon realizar clases a 33 alumnos por sala (*d. s.* = 8), con una mediana de 35 alumnos. En cuanto a su experiencia laboral, los participantes presentaron un promedio de 10 años de experiencia como docentes de aula (*d. s.* = 9) y una mediana de seis años.

Con respecto a su formación inicial, el 97,2% de los encuestados afirmó haber tenido algún curso en el que se abordaron contenidos relativos a la evaluación de aprendizajes; más específicamente, el 56,8% de los participantes señaló haber cursado una asignatura especializada sobre el tema. Posterior a esta formación, el 78,2% de los encuestados reconoce haber participado de algún perfeccionamiento, en el que trataron contenidos relativos a la evaluación de aprendizajes; siendo el curso presencial, la modalidad más frecuente entre los encuestados (52,3%) (detalle de la caracterización de los participantes en Anexo 5).

3.2.2.2. Características del instrumento

El instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes corresponde a un cuestionario individual de administración virtual. Considerando las características del constructo medido, se decidió optar por un instrumento de autorreporte, en tanto este tipo de cuestionario permite recabar información relativa a las disposiciones internas de los sujetos estudiados; por esta razón se realizó una revisión bibliográfica (e.g., Chan, 2009; Van de Mortel, 2008) relativa a los posibles sesgos que se podían generar en este tipo de instrumentos, a fin de evitarlos. El cuestionario se conformó por reactivos de selección de respuesta formato Likert, distribuidos en tres escalas psicológicas provenientes de la Teoría de Conducta Planificada: Componente de intención conductual, Intención conductual y Conducta. La tabla de especificaciones presentada en la Tabla 8 detalla la estructura del instrumento.

Tabla 9: Tabla de especificaciones del instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes, versión pilotaje

Escalas	Subescalas	Posición de reactivos*	Tipo de preguntas	Nivel de medición	Cantidad de reactivos	%
Componentes de la Intención conductual	Actitud cognitiva	4, 28, 29(-), 34, 47, 56, 65, 78	Escala Likert, 4 puntos (Cerrada)	Ordinal	8	10%
	Actitud afectiva	2, 11(-), 13, 21, 35, 68, 70, 80(-)	Escala Likert, 4 puntos (Cerrada)	Ordinal	8	10%
	Norma subjetiva desde los superiores	1, 5, 19, 26, 42, 44, 49, 57	Escala Likert, 4 puntos (Cerrada)	Ordinal	8	10%
	Norma subjetiva desde pares	12, 23, 33(-), 40(-), 43(-), 55(-), 72(-), 73	Escala Likert, 4 puntos (Cerrada)	Ordinal	8	10%
	Control percibido limitado por las competencias	6(-), 37, 50(-), 58(-), 61, 67(-), 75, 77(-)	Escala Likert, 4 puntos (Cerrada)	Ordinal	8	10%
	Control percibido limitado por tiempo	7, 9, 14(-), 16(-), 31, 52, 60, 71	Escala Likert, 4 puntos (Cerrada)	Ordinal	8	10%
Intención conductual		3, 15, 17, 22, 27, 30, 32, 39, 41, 45, 48, 51, 53, 54, 59, 64	Escala Likert, 4 puntos (Cerrada)	Ordinal	16	20%
Conducta		8, 10, 18, 20, 24, 25, 36, 38, 46, 62, 63, 66, 69, 74, 76, 79	Escala Likert, 4 puntos (Cerrada)	Ordinal	16	20%

Nota: *Ítems seguidos de (-) están redactados inversamente; es decir, a mayor puntuación en la respuesta, menor presencia del constructo medido.

La presentación del instrumento fue precedida por un consentimiento informado en el que se dio a conocer a los participantes el objetivo del estudio. Adicionalmente, para caracterizar la muestra con mayor rigurosidad, se incluyó una sección referida principalmente a la formación académica y a los antecedentes laborales de los participantes. Considerando estas secciones más la resolución del instrumento mismo, se estimó un tiempo total de respuesta de 30 minutos (ver Anexo 6).

3.2.2.3. Plan de análisis

Para analizar la información recabada en el pilotaje, se emplearon los programas FACTOR (versión 10.3.01) (Lorenzo-Seva y Ferrando, 2015), JMetrik (versión 3.1.2) (Meyer, 2014) e IBM SPSS Statistics (versión 21.0) (IBM Corporation, 2012). La Tabla 10 detalla los análisis realizados para el pilotaje y los programas empleados en cada uno de ellos.

En primer lugar, se analizaron descriptivamente los datos de la muestra, calculándose puntajes promedio, desviación estándar, mínimo y máximo. Adicionalmente, con objeto de establecer si la distribución de datos se asemejaba a una distribución normal, se emplearon como referentes los estadísticos de simetría y curtosis multivariados (Mardia, 1970) y univariados (Muthén y Kaplan, 1985, 1992).

Luego, se procedió a analizar la consistencia interna del instrumento, calculada a través del coeficiente alfa de Cronbach (1951), además de la correlación ítem-test. Para interpretar estos valores, se emplearon los planteamientos de Hogan (2004) y Murphy y Davidshofer (2005), considerándose desde 0,80 un nivel aceptable y desde 0,90 un nivel elevado de consistencia interna.

Tabla 10: Análisis psicométricos realizados en el pilotaje y programas estadísticos empleados

Análisis psicométricos	Programas estadísticos
Estadísticos descriptivos	FACTOR (versión 10.3.01)
	IBM SPSS Statistics (versión 21.0)
Consistencia interna	IBM SPSS Statistics (versión 21.0)
Capacidad discriminativa	JMetrik (versión 3.1.2)
	IBM SPSS Statistics (versión 21.0)
Análisis factorial	FACTOR (versión 10.3.01)
	IBM SPSS Statistics (versión 21.0)
Análisis de diferencia de grupos	IBM SPSS Statistics (versión 21.0)

Para analizar la capacidad discriminativa de los ítems, en primer lugar, se obtuvo el índice de discriminación basado en grupos extremos⁴ (Anastasi y Urbina, 1998; Cohen y Swerdik, 2000) efectuando el test de U de Mann-Whitney (1947) para muestras independientes. Adicionalmente, se calculó el índice de capacidad discriminativa mediante el programa JMetrik (Meyer, 2014), considerando aceptables valores desde 0,30 y satisfactorios desde 0,40 (Cohen y Swerdlik, 2000; Crocker y Algina, 1986).

Para obtener evidencia con respecto a la validez de la estructura interna del instrumento, se llevó a cabo un Análisis factorial exploratorio y un Análisis factorial confirmatorio del cuestionario completo, además de un Análisis factorial referido a cada una de las escalas teóricas que lo conforman.⁵ Previamente, se revisó el ajuste de los datos a los requerimientos de estos análisis, considerándose un tamaño muestral aceptable desde los 200 participantes, con un ratio mínimo de 5 participantes por ítem (Comrey y Lee, 1992; Gorsuch, 1983). Además, se evaluó estadísticamente la intercorrelación de los ítems a través del test de esfericidad de Bartlett y la medida de adecuación muestral de Kaiser-Mayer-Olkin [KMO] (Everitt y Wykes, 2001; Hair, Anderson, Tatham y Black, 1999).

⁴ Para conformar los grupos extremos, se consideró a los participantes con menor puntaje total (hasta el percentil 33) y con mayor puntaje total (desde el percentil 67) (Anastasi y Urbina, 1998).

⁵ Cabe señalar que, de contar con un tamaño muestral mayor, se habría realizado un análisis factorial confirmatorio con una muestra distinta a la utilizada en el análisis factorial exploratorio, dividiendo aleatoriamente a la muestra en dos mitades; no obstante, considerando el tamaño muestral de este estudio, se empleó la misma muestra para correr ambos análisis, sin descuidar la rigurosidad de los mismos y reconociendo en ello una de sus limitaciones.

Para la extracción de factores se optó por el método factorial de Ejes principales. A diferencia del método de Componentes principales, el método factorial de Ejes principales supone la existencia de un modelo causal subyacente en los datos analizados (Costello y Osborne, 2005; Tabachnick y Fidell, 2001); por otra parte, se desestimó el uso del método factorial de Máxima verosimilitud, debido a que sólo es pertinente en datos que presentan una distribución normal, no siendo éste el caso (Costello y Osborne, 2005). Para decidir la cantidad de factores a extraer, se consideró el Análisis paralelo (Horn, 1965; Timmerman y Lorenzo-Seva, 2011) y el Criterio gráfico (Cattell, 1966). El Análisis paralelo destaca por su precisión y por su fundamento teórico y estadístico; sin embargo, aún no se ha generalizado su uso (Ledesma y Valero-Mora, 2007). Por el contrario, el Criterio gráfico es uno de los más usados en la actualidad y aun cuando posee un alto componente subjetivo, se ha verificado la confiabilidad de sus soluciones (Kline, 2000; Pérez y Medrano, 2010). De este modo, la utilización de ambos referentes permite una decisión más robusta. Se descartó el uso del Criterio de Kaiser (1960), dado que, además de presentar soluciones alejadas a las sugeridas por los criterios anteriores, conduce a la extracción una cantidad de factores superior a la necesaria, especialmente en instrumentos con más de 50 ítems (Pérez y Medrano, 2010). En cuanto a la rotación, se decidió efectuar una rotación oblicua Oblimin Directo, ya que es la más adecuada en constructos psicológicos, donde suele existir una mayor correlación entre las variables (Gorsuch, 1983). Finalmente, para la distribución de los reactivos en los factores resultantes, se consideraron correlaciones o cargas factoriales absolutas, entre un ítem y su factor subyacente de al menos 0,35, sin existir una correlación superior a 0,30 de ese reactivo con otro factor (Pérez y Medrano, 2010).

Con objeto de robustecer la evidencia a favor de la validez del instrumento, se realizaron pruebas estadísticas no paramétricas basadas en el rango, las que permitieron analizar diferencias entre grupos definidos a partir de características sociodemográficas. Se optó por pruebas no paramétricas dado que la distribución de los datos no se asemejó a una distribución normal, presentando estadísticos de asimetría y curtosis que avalaron esta decisión. En el caso de la variable género, se empleó el test de U de Mann-Whitney (1947), dado que dicha variable divide a la muestra en dos grupos independientes: femenino y masculino. Con respecto a la variable geográfica, considerando que las regiones de Chile a las que pertenecen los participantes dividen a la muestra en más de dos grupos independientes, se decidió emplear el test de Kruskal Wallis (1952). Ambos tests permitieron analizar diferencias de grupos, reconociendo la naturaleza ordinal de la escala Likert empleada, radicando su distinción en la cantidad de grupos que se originan a partir de la variable sociodemográfica estudiada (Rivas-Ruiz, Moreno-Palacios y Talaveraa, 2013). A partir de ello se detectó si existían diferencias significativas en las respuestas obtenidas entre estos grupos para cada uno de los reactivos, utilizando un nivel de confianza de un 95% ($p < 0,05$). Estos análisis no constituyen en sí mismos evidencias que permitan aseverar o refutar la presencia de sesgo, entendiéndose como tal la validez diferencial de una interpretación de la puntuación de un instrumento para un grupo particular; sólo se puede hablar de sesgo, en la medida que ello no corresponda a una diferencia real en el rasgo subyacente que la prueba trata de medir (Hogan, 2004; Martínez, 1995). No obstante, estos análisis sí permiten dar cuenta de un posible funcionamiento diferencial de los ítems que presenten diferencias significativas para los grupos analizados, pues a priori las características sociodemográficas no se relacionan con el constructo que se persigue medir, resultando importante considerar esta información al momento de validar el cuestionario en su globalidad.⁶

⁶ Cabe recalcar el estatus de *posible funcionamiento diferencial* y no de *funcionamiento diferencial* a secas, dado que para ello se requerirían análisis estadísticos complementarios, siendo ésta una limitación de este estudio.

3.3. Preparación de instrumento final

Luego de realizar los análisis psicométricos, se procedió a hacer una nueva selección de ítems basada en los resultados antes obtenidos con objeto de constituir una nueva versión del instrumento, resguardando la validez de sus reactivos. De este modo, se utilizó como criterio primordial la estructura factorial, escogiendo a partir de ello los ítems que permitiesen alcanzar una consistencia interna más alta, que discriminasen mejor entre los diferentes respondentes y que presentasen menores diferencias entre grupos acorde al género y a la región de los participantes.

Una vez efectuada dicha selección de ítems, se analizaron las propiedades psicométricas de esta propuesta, replicando los análisis de la validación, esta vez focalizándose en los reactivos escogidos. De este modo, la propuesta finalmente presentada contó con robusta evidencia a favor de su validez, además de coeficientes adecuados de confiabilidad.

IV. RESULTADOS

4.1. Estadísticos descriptivos

Con respecto a los estadísticos descriptivos del instrumento, se obtuvo una media de 253 puntos ($d.s. = 23$). El puntaje mínimo alcanzado fue 140 puntos y el máximo 308 puntos, de un total de 400 puntos (ver Anexo 7).

Por otra parte, al analizar la distribución de los datos de la muestra, no se rechazó la hipótesis de que la distribución fuese simétrica ($\chi^2_{88560} = 1496,911$; $p > 0,05$), aunque sí se rechazó que fuese mesocúrtica ($\chi^2_{88560} = 7239,535$; $p < 0,05$); por tanto, no es posible asumir una distribución normal multivariada de los datos. Respecto de los descriptivos univariados, en más de un tercio de los ítems, los índices de simetría y curtosis se encontraron fuera de un rango de $\pm 1,5$ (ver Anexo 8).

4.2. Consistencia interna

El instrumento completo presentó un nivel elevado de consistencia interna, con un coeficiente alfa de Cronbach de 0,915. A partir del análisis de correlación ítem-test, se obtuvo que de eliminarse los ítems 2, 12, 14, 16, 29, 33, 40, 42, 44, 55, 67, 72, 77 y 80, la consistencia interna podría aumentar a 0,933, siendo importante considerar su revisión (ver Anexo 9).

4.3. Capacidad discriminativa de los ítems

De acuerdo al índice de discriminación basado en grupos extremos, la mayor parte de los reactivos presentó diferencias significativas entre las distribuciones de puntajes de estos grupos ($p < 0,05$), salvo en el caso de los ítems 16, 29, 33, 42 y 72. Dado que no se obtuvo evidencia a favor de la capacidad discriminativa de estos reactivos, surge la necesidad de revisar su formulación (ver Anexo 10).

Adicionalmente, se calculó el índice de capacidad discriminativa. Este análisis reveló una capacidad discriminativa leve ($0,20 \leq D < 0,30$) en los ítems 1, 6, 11, 23, 34, 43, 50, 58, 60, 67 y 71. Por otra parte, los ítems 2, 12, 14, 16, 29, 33, 40, 42, 44, 55, 72, 77 y 80 presentaron capacidad discriminativa cercana a cero ($D < 0,20$), considerándose su revisión; en especial en el caso de los ítems 29 y 80, pues estarían discriminando inversamente ($D < 0$) (ver Anexo 11). La curva característica de estos ítems (Fig. 5), representación gráfica de su comportamiento, corroboró lo anteriormente señalado.

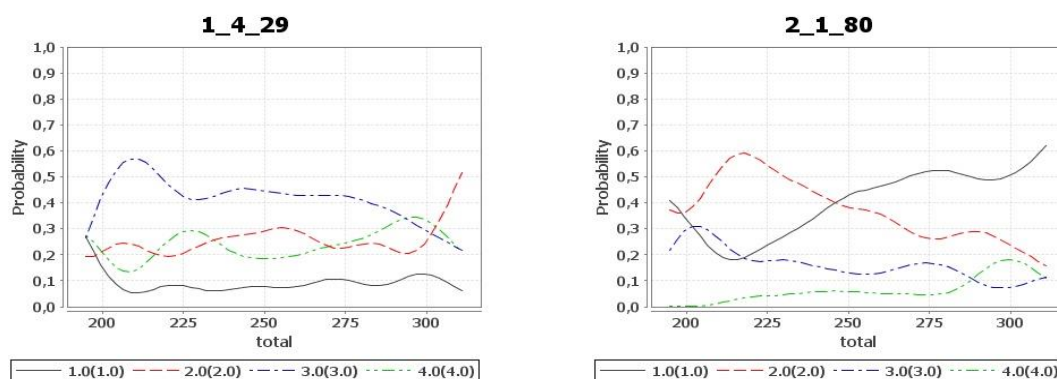


Fig. 5: Curva característica de ítems que discriminan inversamente (29 y 80).

4.4. Análisis factorial

4.4.1. Análisis factorial del instrumento completo

La muestra, conformada por 509 docentes y con un ratio de 6,363 participantes por variable⁷, presentó un tamaño adecuado para efectuar un Análisis factorial. Por otra parte, el test de esfericidad de Bartlett ($\chi^2_{3160} = 16352,163$; $p < 0,05$) y la medida de adecuación muestral de Kaiser-Mayer-Olkin [KMO] (0,905), proporcionaron evidencia satisfactoria que verificó la pertinencia del uso de este análisis.

⁷ Cociente entre los 509 participantes y las 80 afirmaciones relativas al instrumento completo.

Debido a los datos no presentaron una distribución normal multivariada, la extracción escogida fue el método factorial de Ejes Principales con una rotación oblicua Oblimin Directo. De acuerdo a lo sugerido por el Criterio gráfico (Fig. 6) y por el Análisis paralelo (ver Anexo 12) se decidió extraer seis factores, los que explicaron el 38,53% de la variabilidad total de las respuestas al cuestionario (Tabla 11).

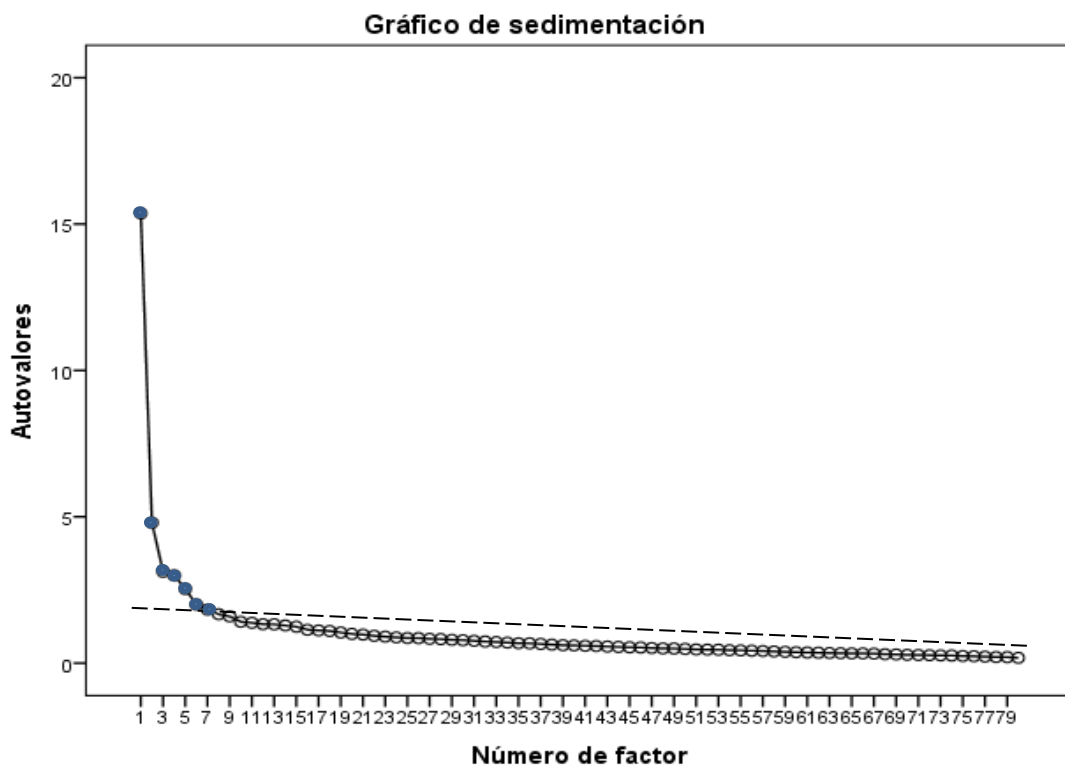


Fig. 6: Gráfico de sedimentación de Análisis factorial del instrumento completo con seis factores.

Tabla 11: Varianza explicada basada en los autovalores del Análisis factorial exploratorio con seis factores

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	15,365	19,206	19,206	14,758	18,448	18,448
2	4,798	5,997	25,203	4,209	5,261	23,709
3	3,124	3,905	29,109	2,498	3,122	26,831
4	2,997	3,747	32,855	2,329	2,911	29,742
5	2,537	3,172	36,027	1,862	2,327	32,069
6	2,002	2,503	38,530	1,361	1,701	33,770

En un primer Análisis factorial del instrumento completo, los reactivos 39 y 64 presentaron cargas absolutas superiores a 0,30 en dos factores, sugiriéndose por ello su revisión a fin de suprimir información redundante. Por otra parte, los ítems 2, 8, 10, 11, 12, 20, 29, 34, 36, 37, 38, 42, 44, 46, 49, 77 y 80 presentaron cargas absolutas inferiores a 0,35 en todos los factores, sugiriéndose también su revisión, pues estos ítems podrían no caracterizar lo suficiente a los factores que componen el constructo. Adicionalmente, a partir de la revisión del contenido de los ítems por factor, se identificó que los ítems 13, 19, 24, 27, 32, y 76 no se ajustaban a la caracterización teórica de los factores en los que se agrupaban; por esta razón, se decidió revisar su formulación (ver Anexo 13).

En un segundo Análisis factorial, se agruparon los 55 reactivos restantes en seis factores, explicando el 43,64% de la varianza total. Los datos cumplieron con los requisitos de pertinencia del análisis, de acuerdo a los estadísticos del test de esfericidad de Bartlett ($\chi^2_{1485} = 10295,360$; $p < 0,05$) y de medida de adecuación muestral de KMO (0,894). En este análisis, los ítems 61 y 75 presentaron cargas absolutas inferiores a 0,35 en todos los factores, siendo relevante revisar su formulación (ver Anexo 14).

Al replicar el análisis factorial sin estos reactivos, explicando el 43,828% de la varianza total, se obtuvieron estadísticos asociados al test de esfericidad de Bartlett ($\chi^2_{1378} = 9754,174$; $p < 0,05$) y a la medida de adecuación muestral de KMO (0,889) que corroboraron su pertinencia. En este análisis, el ítem 73 presentó una carga absoluta inferior a 0,35 en todos los factores, sugiriéndose su revisión (ver Anexo 15).

Finalmente, se realizó un Análisis factorial de seis factores con los 52 ítems restantes, explicando el 44,03% de la varianza total. Los estadísticos asociados al test de esfericidad de Bartlett ($\chi^2_{1326} = 9523,728$; $p < 0,05$) y a la medida de adecuación muestral de KMO (0,892) dieron cuenta de la pertinencia de este análisis. Los 52 ítems presentaron cargas absolutas superiores a 0,35 en un factor, además de coherencia con la revisión bibliográfica previa (ver Anexo 16). La caracterización de los factores obtenidos se presenta en la Tabla 12, interpretándose en relación a fundamentos teóricos relacionados a la evaluación de aprendizajes y a la Teoría de Conducta Planificada.

Tabla 12: Caracterización de factores extraídos del Análisis Factorial del instrumento completo

Factor	Características teóricas	Ítems
1	TCP: Actitud (cognitiva y afectiva), Intención conductual y Conducta. Modelo de Evaluación de aprendizajes: Sistémico.	3, 4, 15, 17, 18, 21, 22, 25, 28, 30, 35, 41, 45, 47, 48, 51, 53, 54, 59, 62, 66, 68, 69, 74, 78.
2	TCP: Control percibido referido a tiempo.	7, 9, 14, 16, 31, 52, 60, 71.
3	TCP: Norma subjetiva (negativa), referida a otros docentes.	33, 40, 43, 55, 72.
4	TCP: Actitud (cognitiva y afectiva) y Conducta. Modelo de Evaluación de aprendizajes: Comunicativo.	56, 63, 65, 70, 79.
5	TCP: Norma subjetiva (positiva), referida a superiores y a otros docentes.	1, 5, 23, 26, 57.
6	TCP: Control percibido referido a competencias.	6, 50, 58, 67.

En la tabla anterior, es posible observar que los factores están caracterizados principalmente de acuerdo a las variables de la Teoría de Conducta Planificada. No obstante, en el Factor 1 se agrupó la Actitud, la Intención conductual y la Conducta, al igual que en el Factor 4, la Actitud y la Conducta; por ello, resultó pertinente realizar adicionalmente un Análisis factorial por escala que respete la estructura teórica establecida por el modelo, separando los Componentes de la intención conductual de las otras escalas. Por otra parte, de la estructura teórica relativa a la evaluación de aprendizajes, es posible observar que sólo el Factor 4 aludió a las innovaciones que el Modelo comunicativo adicionó a su antecesor, con reactivos que apuntaban a la incorporación activa del alumnado en el proceso evaluativo; pese a ello, los demás reactivos aun correspondían al Modelo comunicativo, en tanto aludían a características que este enfoque conservó del Modelo sistémico. En consecuencia, de las categorías inicialmente establecidas referidas a las prácticas evaluativas de los docentes, éstas se dividieron en dos: prácticas evaluativas que incorporan activamente al alumnado en el proceso evaluativo (Factor 4) y prácticas evaluativas que abordan la evaluación como proceso (Factor 1, Factor 2, Factor 3, Factor 5 y Factor 6), agrupando en este último tipo los reactivos relacionados a la obtención de información edumétricamente válida y confiable, a la retroalimentación orientada hacia la mejora en los aprendizajes y a la responsabilización docente dentro del proceso evaluativo.

4.4.2. Análisis factorial del instrumento por escalas

Considerando las escalas predefinidas de la Teoría de Conducta Planificada, se realizó adicionalmente un Análisis factorial con esta estructura. Los datos de las escalas Componentes de la intención conductual, Intención conductual y Conducta presentaron características que corroboraron la pertinencia del análisis, en términos de tamaño muestral y de estadísticos asociados al test de esfericidad de Bartlett y a la medida de adecuación muestral de KMO (Tabla 13).

Tabla 13: Características del Análisis factorial del instrumento por escalas que corroboran la pertinencia del análisis

Escala	Ratio participantes por ítem*	Test de esfericidad test de esfericidad de Bartlett	Medida de adecuación muestral de KMO
Componentes de la intención conductual	10,604	$\chi^2_{1128} = 6179,911; p < 0,05$	0,823
Intención conductual	31,813	$\chi^2_{561} = 2813,842; p < 0,05$	0,932
Conducta	31,813	$\chi^2_{528} = 2506,843; p < 0,05$	0,905

Nota: *Este ratio corresponde al cociente entre los 509 participantes y la cantidad de afirmaciones relativas a cada escala.

Al analizar la distribución de los datos de la muestra, no fue posible asumir una distribución normal multivariada de los mismos, pues si bien no se rechazó la hipótesis de que la distribución fuese simétrica, sí se rechazó que fuese mesocúrtica. Respecto de los descriptivos univariados, los índices de simetría y curtosis se encontraron en más de un tercio de los ítems fuera de un rango de $\pm 1,5$ (Tabla 14).

Tabla 14: Estadísticos descriptivos multivariados y univariados por escala

Escala	Test de Mardia referido a la distribución multivariada de los datos		Porcentaje de ítems que presentan índices univariados fuera de un rango de $\pm 1,5$	
	Simetría	Curtosis	Simetría	Curtosis
Componentes de la intención conductual	$\chi^2_{19600} = 361,356; p > 0,05$	$\chi^2_{19600} = 2629,305; p < 0,05$	16,67% (8 de 48 ítems)	20,83% (10 de 48 ítems)
Intención conductual	$\chi^2_{816} = 79,999; p > 0,05$	$\chi^2_{816} = 465,034; p < 0,05$	87,50% (14 de 16 ítems)	87,50% (14 de 16 ítems)
Conducta	$\chi^2_{816} = 42,085; p > 0,05$	$\chi^2_{816} = 369,136; p < 0,05$	37,50% (6 de 16 ítems)	56,15% (9 de 16 ítems)

Considerando la distribución de los datos, dados los estadísticos multivariados y univariados, se decidió optar por el método factorial de Ejes Principales con una rotación oblicua Oblimin Directo. A partir del Análisis paralelo (ver Anexo 17) y del Criterio gráfico (Fig. 7), se decidió extraer seis factores para la escala Componentes de la intención conductual, uno para la escala Intención conductual y dos para la escala Conducta; factores que explicaron el 38,97%, el 39,41% y el 44,31% de la variabilidad total de las respuestas, respectivamente.

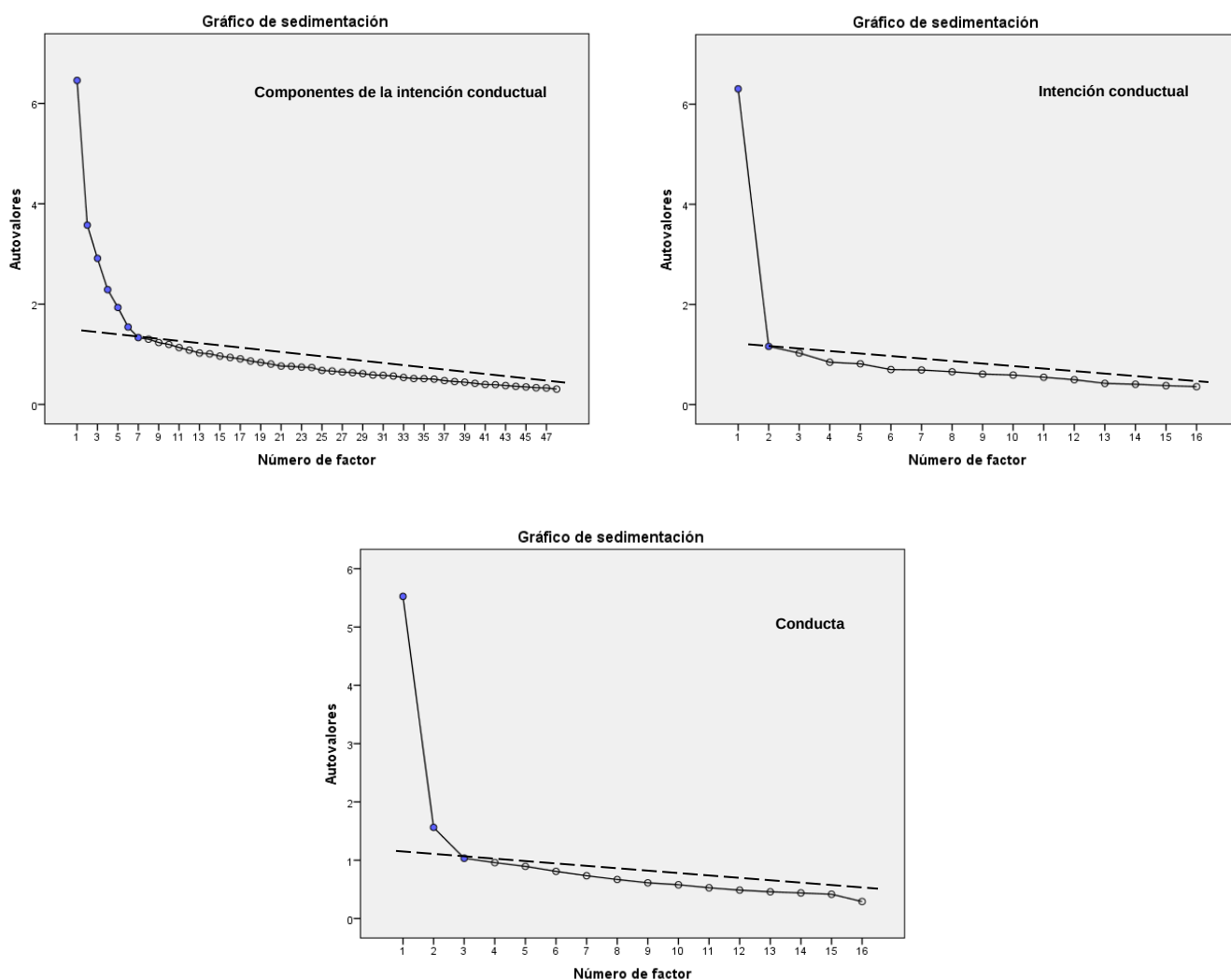


Fig. 7: Gráficos de sedimentación por escala.

En un primer Análisis factorial de los 48 ítems que conformaban la escala Componentes de la intención conductual, el ítem 73 presentó una carga absoluta superior a 0,30 en dos factores, sugiriéndose su revisión a fin de evitar información redundante entre factores. Por otra parte, los ítems 11, 12, 23, 29, 34, 37, 42, 44, 49, 77 y 80 presentaron una carga factorial absoluta inferior a 0,35 en todos los factores, siendo ello razón para revisar su formulación dada la insuficiente evidencia empírica a favor de la caracterización de estos ítems respecto de los factores que conforman el constructo. Respecto de la revisión de contenido de los reactivos, los ítems 13 y 19 no se ajustaron conceptualmente a los factores en los que se agruparon, siendo pertinente revisar su formulación (ver Anexo 18).

Con las 34 afirmaciones restantes, se realizó un segundo Análisis factorial de seis factores los cuales explicaron el 47,017% de la varianza total de las respuestas. Se obtuvieron estadísticos asociados al test de esfericidad de Bartlett ($\chi^2_{561} = 4360,031$; $p < 0,05$) y a la medida de adecuación muestral de KMO (0,831), los que corroboraron su pertinencia. En este análisis, sólo el ítem 68 presentó una carga absoluta superior a 0,30 en dos factores, sugiriéndose por ello su revisión a fin de evitar información redundante entre factores (ver Anexo 19).

En consecuencia, para un tercer Análisis factorial se conservaron 33 ítems, los que explicaron un 47,36% de la variabilidad de las respuestas. Se corroboró la pertinencia del análisis mediante los estadísticos asociados al test de esfericidad de Bartlett ($\chi^2_{528} = 4171,756$; $p < 0,05$) y a la medida de adecuación muestral de KMO (0,826). En este análisis, los 33 ítems presentaron una carga factorial absoluta superior a 0,35 sólo en un factor, sin presentar cargas factoriales absolutas de más de 0,30 en el resto de los factores; además, resultaron ser coherentes con el contenido de los reactivos que conformaban sus respectivos factores (ver Anexo 20). La Tabla 15 presenta la caracterización de los factores resultantes a partir de los resultados empíricos del análisis anterior y a la revisión bibliográfica realizada previamente.

Tabla 15: Caracterización de factores extraídos del Análisis Factorial, escala Componentes de la intención conductual

Factor	Características teóricas	Ítems
1	TCP: Control percibido referido a tiempo.	7, 9, 14, 16, 31, 52, 60, 71.
2	TCP: Actitud cognitiva y afectiva. Modelo de Evaluación de aprendizajes: Sistémico.	2, 4, 21, 28, 35, 47, 78.
3	TCP: Norma subjetiva (negativa), referida a otros docentes.	33, 40, 43, 55, 72.
4	TCP: Control percibido referido a competencias.	6, 50, 58, 61, 67, 75.
5	TCP: Actitud (cognitiva y afectiva) y conducta. Modelo de Evaluación de aprendizajes: Comunicativo.	56, 65, 70.
6	TCP: Norma subjetiva (positiva), referida a superiores	1, 5, 26, 57.

Posteriormente, se llevó a cabo el Análisis factorial de la escala Intención conductual con un factor. Esta escala está conformada por 16 ítems: 3, 15, 17, 22, 27, 30, 32, 39, 41, 45, 48, 51, 53, 54, 59 y 64. En el análisis factorial de dicha escala no se obtuvo reactivos con una correlación absoluta inferior a 0,35 en su factor único ni inconsistencias con la teoría subyacente (ver Anexo 21).

Finalmente, se procedió a efectuar el Análisis factorial con dos factores en la escala Conducta, conformada por 16 ítems. En este análisis, los ítems 24, 36 y 76 presentaron cargas factoriales absolutas sobre 0,30 en ambos factores, sugiriéndose su revisión a fin de evitar redundancia en la información provista por éstos. Por otro lado, los ítems 8, 20 y 46 presentaron una carga factorial inferior a 0,35 en ambos factores, siendo pertinente también su revisión, dado que no evidenciarían suficientemente la caracterización de los factores que conforman el constructo estudiado (ver Anexo 22).

Con los 10 ítems restantes de esta escala se efectuó un segundo Análisis factorial considerando dos factores, explicando el 52,35% de la variabilidad de las respuestas. Se corroboró nuevamente la pertinencia del análisis mediante los estadísticos asociados al test de esfericidad de Bartlett ($\chi^2_{45} = 1376,776$; $p < 0,05$) y a la medida de adecuación muestral de KMO (0,836). En este análisis, los 10 ítems presentaron una carga factorial absoluta superior a 0,35 sólo en un factor, sin presentar cargas factoriales absolutas de más de 0,30 en el otro factor; además, presentaron una consistencia teórica con el resto de los reactivos agrupados en sus respectivos factores (ver Anexo 23). La Tabla 16 presenta la caracterización de los factores resultantes a partir de los resultados empíricos del análisis anterior y a la revisión bibliográfica realizada previamente.

En el Análisis factorial efectuado en este apartado, se consideró a priori la distribución teórica de las escalas de la Teoría de Conducta Planificada. En el caso de la escala Componentes de la intención conductual, se obtuvo una distribución similar a la hallada al analizar el instrumento completo; en consecuencia, las subescalas surgidas a partir de ella distinguieron la Actitud de la Norma subjetiva y del Control percibido. Tanto la Norma subjetiva como el Control percibido presentaron empíricamente dos subescalas cada uno (Factor 3 y Factor 6, Factor 2 y Factor 4, respectivamente), coincidiendo con la subdivisión establecida durante la construcción. Por su parte, la Actitud dio origen a dos subescalas, donde una de ellas apuntó a las innovaciones del Modelo comunicativo (Factor 2) y la otra a los aportes que este enfoque mantuvo de su antecesor (Factor 5).

Tabla 16: Caracterización de factores extraídos del Análisis Factorial, escala Conducta

Factor	Características teóricas	Ítems
1	TCP: Conducta. Modelo de Evaluación de aprendizajes: Sistémico.	18, 25, 38, 62, 66, 69, 74
2	TCP: Conducta. Modelo de Evaluación de aprendizajes: Comunicativo.	10, 63, 79

Respecto de la escala de Intención conductual, se obtuvo un único factor en el que no se presentó la división antes señalada entre los diferentes enfoques de evaluación educativa. Sin embargo, en el caso de la escala Conducta, sí fue posible distinguirla, pues el Factor 2 agrupó los ítems asociados a las innovaciones del Modelo comunicativo, mientras que el Factor 1 aludió a los aportes que este enfoque conservó de su antecesor, en términos la obtención de información edumétricamente válida y confiable, a la retroalimentación orientada hacia la mejora en los aprendizajes y a la responsabilización docente dentro del proceso evaluativo.

4.5. Análisis de diferencias entre grupos

Se realizó el test no paramétrico de U de Mann-Withney para comparar dos muestras independientes, definidas de acuerdo con la variable género. A partir de esta prueba, se encontraron diferencias significativas en las respuestas obtenidas entre grupos para los ítems 2, 13, 24, 25, 38, 48, 50, 53, 59, con un nivel de confianza de un 95% ($p < 0,05$) (ver Anexo 24).

Por otra parte, se aplicó test no paramétrico de Kruskal Wallis para comparar más de dos muestras independientes, definidas acorde a la región del país en la que se desempeñan los participantes. La prueba dio cuenta de diferencias significativas en las respuesta obtenidas entre grupos para los ítems 6, 28, 35 y 53, con un nivel de confianza de un 95% ($p < 0,05$) (ver Anexo 25).

De este modo, se consideró la revisión de los reactivos antes mencionados, a fin de reducir la potencial existencia de sesgo producto de características demográficas que no guardan directa relación con el constructo que el instrumento intenta medir.

4.6. Análisis consolidado de reactivos

La información obtenida fue consolidada y revisada considerando la consistencia interna, su capacidad discriminativa, su agrupación factorial en el instrumento y por escalas, diferencias por género y diferencias por región, con el fin de elegir de manera rigurosa los reactivos que formarían parte de una versión post pilotaje del cuestionario (ver Anexo 26). De este modo, la versión propuesta se conformó de 30 ítems distribuidos en tres escalas: Componentes de intención conductual (18 ítems), Intención conductual (6 ítems) y Conducta (6 ítems) (ver Anexo 27).

Considerando los datos empíricos de los ítems de esta versión, la media del instrumento fue 96 puntos ($d.s. = 11$), con un mínimo alcanzado de 48 puntos y un máximo alcanzado de 120 puntos, coincidiendo este último valor con el máximo teórico.

El índice alfa de Cronbach correspondió a 0,868 para el instrumento completo, y de 0,750, 0,824 y 0,828 para las escalas Componentes de la intención conductual, Intención conductual y Conducta, respectivamente (detalle respecto de la variación de la consistencia interna del instrumento y sus escalas si se eliminase una pregunta en Anexo 28). Por otra parte, no se obtuvieron ítems que discriminasen inversamente a los respondentes; por el contrario, 23 ítems de los 30 ítems (77%) presentaron un índice de capacidad discriminativa sobre 0,3 (ver Anexo 29).

Para evaluar la pertinencia del Análisis factorial respecto del instrumento propuesto, se calcularon indicadores que confirmaron la idoneidad de efectuar dicho análisis (Tabla 17). La cantidad de factores a extraer en cada escala, fue determinada por el Criterio gráfico basado en el Gráfico de sedimentación por escala (Fig. 8). En la escala Componentes de la intención conductual, se extrajeron seis factores, los que explicaron el 60,97% de la varianza total de la escala. Por su parte, se extrajo un factor tanto en la escala Intención conductual como en la escala Conducta, explicando el 53,49% y el 54,31% de la varianza total de las respuestas de la escala, respectivamente (ver Anexo 30).

Tabla 17: Características del Análisis factorial del instrumento por escalas, versión post pilotaje, que corroboran la pertinencia del análisis

Escala	Ratio participantes por ítem*	Test de esfericidad test de esfericidad de Bartlett	Medida de adecuación muestral de KMO
Componentes de la intención conductual	28,278	$\chi^2_{153} = 1927,014; p < 0,05$	0,770
Intención conductual	84,833	$\chi^2_{15} = 960,596; p < 0,05$	0,855
Conducta	84,833	$\chi^2_{15} = 1048,657; p < 0,05$	0,833

Nota: *Este ratio corresponde al cociente entre los 509 participantes y la cantidad de afirmaciones relativas a cada escala.

El Análisis factorial por escala aportó evidencia a favor de la validez de constructo, coincidiendo la estructura factorial con la estructura planteada por la teoría subyacente. La tabla de especificaciones presentada en la Tabla 18 da cuenta de la caracterización de las escalas y subescalas que conformaron la versión post pilotaje del instrumento, respetando la estructura factorial obtenida empíricamente y significándola de acuerdo a los constructos teóricos.

No se efectuó nuevamente el análisis de diferencia entre grupos, puesto que la selección de reactivos efectuada no consideró ítems que presentasen diferencias significativas asociadas a género y/o a región.

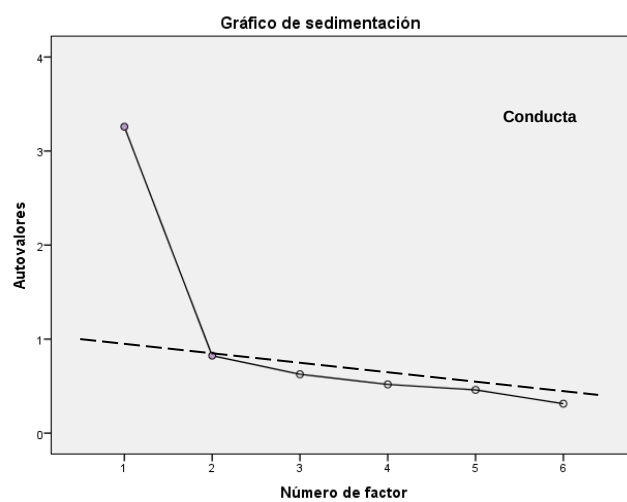
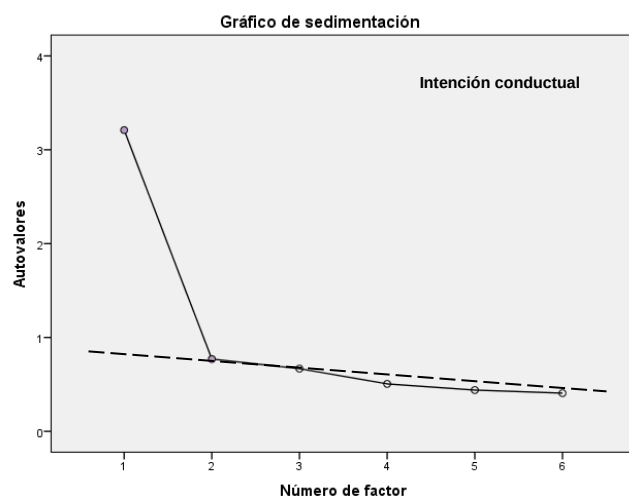
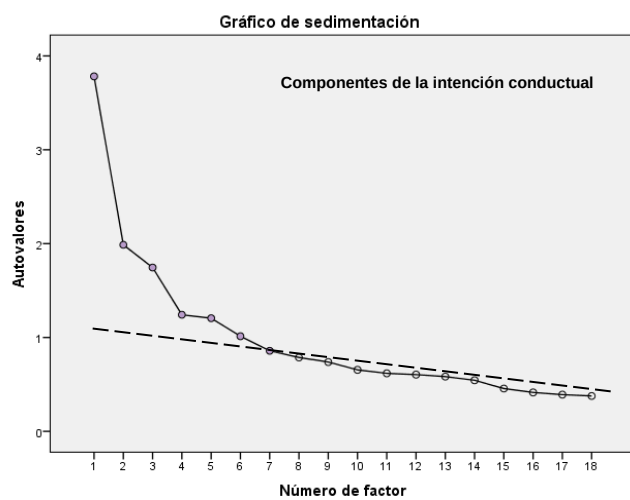


Fig. 8: Gráfico de sedimentación por escala, versión post pilotaje

Tabla 18: Tabla de especificaciones del instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes, versión post pilotaje

Escalas	Subescalas	Posición de reactivos*	Tipo de preguntas	Nivel de medición	Cantidad de reactivos	%
Componentes de la Intención conductual	Actitud cognitiva y afectiva hacia la evaluación de aprendizajes sistémica	4, 21, 78	Escala Likert, 4 puntos (Cerrada)	Ordinal	3	10%
	Actitud cognitiva y afectiva hacia la evaluación de aprendizajes comunicativa	56, 65, 70	Escala Likert, 4 puntos (Cerrada)	Ordinal	3	10%
	Norma subjetiva desde los superiores	5, 26, 57	Escala Likert, 4 puntos (Cerrada)	Ordinal	3	10%
	Norma subjetiva desde pares	40(-), 43(-), 55(-)	Escala Likert, 4 puntos (Cerrada)	Ordinal	3	10%
	Control percibido limitado por las competencias	58(-), 61, 75	Escala Likert, 4 puntos (Cerrada)	Ordinal	3	10%
	Control percibido limitado por tiempo	7, 60, 71	Escala Likert, 4 puntos (Cerrada)	Ordinal	3	10%
Intención conductual		17, 22, 27, 32, 41, 54	Escala Likert, 4 puntos (Cerrada)	Ordinal	6	20%
Conducta		10, 24, 36, 63, 76, 79	Escala Likert, 4 puntos (Cerrada)	Ordinal	6	20%

Nota: *Ítems seguidos de (-) están redactados inversamente; es decir, a mayor puntuación en la respuesta, menor presencia del constructo medido.

En resumen, se comenzó con una versión del instrumento de 100 ítems, los cuales a partir del pretest cognitivo se redujeron a 80. Tras efectuar los análisis psicométricos relativos a su consistencia interna, capacidad discriminativa de los reactivos, estructura factorial y de diferencias entre grupos, se seleccionaron 30 ítems organizados en las tres escalas que conforman la Teoría de Conducta Planificada: Componentes de la intención conductual (18 ítems), Intención conductual (6 ítems) y Conducta (6 ítems). La consistencia interna del instrumento, calculada mediante el índice alfa de Cronbach, fue de 0,868, presentando además una estructura empírica que en el caso de la escala Componentes de intención conductual ($\alpha = 0,750$) se compuso de seis factores, los cuales se significaron en base a la teoría psicológica subyacente y a la revisión bibliográfica acerca de la evaluación educativa: Actitud cognitiva y afectiva hacia la evaluación de aprendizajes sistémica, Actitud cognitiva y afectiva hacia la evaluación de aprendizajes comunicativa, Norma subjetiva desde los superiores, Norma subjetiva desde pares, Control percibido limitado por las competencias, Control percibido limitado por tiempo. Por otro lado, las escalas Intención conductual ($\alpha = 0,824$) y Conducta ($\alpha = 0,828$) se conformaron por factores únicos.

V. DISCUSIÓN

Dentro del proceso educativo, la evaluación de aprendizajes tiene un rol central, puesto que permite reorientar las prácticas pedagógicas del profesorado y las estrategias de los estudiantes, en función de las dificultades y logros que éstos van evidenciando a lo largo de su aprendizaje (Airasian y Russell, 2008; Barkley y Major, 2015; DeLuca et al., 2015; Sanmartí, 2007; Stiggins y Chappuis, 2012).

Dada la importancia de la evaluación de aprendizajes en el proceso educativo, en Chile se han impulsado diferentes iniciativas orientadas al desarrollo de competencias evaluativas en el profesorado (e.g., MINEDUC, 2009a, 2009b). Pese a ello, este aspecto ha sido sistemáticamente uno de los más deficientes dentro de su desempeño (Herrada y Zapata, 2015; MINEDUC, 2008, 2011c, 2014, 2015); situación que se replica en otras partes del mundo (Baird, 2010; DeLuca, 2012; DeLuca et al., 2013; DeLuca y Klinger, 2010; McGee y Colby, 2014; Popham, 2011). De esta manera, resulta relevante estudiar cuáles son los condicionantes que afectan positiva o negativamente la concreción de las prácticas evaluativas de los docentes; siendo, en particular, su disposición interna hacia la evaluación de aprendizajes, el foco de este proyecto. De acuerdo a De la Torre (1998, 2002), es este factor intrínseco del docente el que filtra el contexto socioeducativo en el que éste se desempeña.

En consecuencia, este estudio se focalizó en las disposiciones internas de los docentes, medidas a partir de sus intenciones conductuales hacia prácticas evaluativas que se desprenden del Modelo comunicativo de Quinquer (2000). Dado que la evaluación de aprendizajes se inserta dentro de un espacio social, resultó pertinente abordar este estudio desde la psicología social; particularmente, desde la Teoría de Conducta Planificada (Ajzen, 1991; 2002).

El objetivo de este estudio fue construir y validar un instrumento de medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes, dirigido a profesores de Enseñanza Básica y Media del sistema escolar chileno, que estén ejerciendo docencia en aula. Para cumplir este propósito, se construyeron inicialmente 648 ítems, basados en la revisión de literatura especializada en evaluación de aprendizajes y en la Teoría de Conducta Planificada. Con el fin de validar el contenido de los reactivos, éstos fueron sometidos a un juicio de expertos y a un pretest cognitivo. Considerando ambos procesos, se elaboró una primera versión del cuestionario, conformada por 80 ítems. Esta versión del instrumento fue aplicada masivamente a una muestra representativa de la población de 509 docentes. A partir de este pilotaje, se efectuaron una serie de análisis que permitieron validar psicométricamente el cuestionario.

Respecto del análisis de correlación ítem-test, se identificaron 14 reactivos que, de eliminarse, podrían aumentar la consistencia interna del instrumento. Dado que el índice alfa de Cronbach es un indicador de homogeneidad de los reactivos, este resultado dio cuenta de que los ítems anteriores estaban midiendo algo diferente al resto de las afirmaciones que conformaron el instrumento. Asimismo, 13 de estos 14 ítems demostraron tener también una capacidad discriminativa cercana a cero, siendo en dos de ellos negativa. En este caso, dichos reactivos demostraron no diferenciar adecuadamente entre los participantes con alta presencia del rasgo medido, de los participantes con baja presencia del mismo. Al revisar la formulación de las afirmaciones anteriormente comentadas, 10 de ellas estaban formuladas de manera inversa, es decir, su puntuación denotaba ausencia del rasgo medido. De acuerdo a DiStefano y Motl (2006), los ítems redactados de forma negativa podrían tener un efecto de método sobre el estilo de respuesta de los encuestados, que no necesariamente está asociado al constructo medido. Considerando lo anterior, para la selección de los reactivos para la versión post pilotaje, se evitó la elección de dichas afirmaciones.

El Análisis factorial del instrumento completo y por escala, aportó evidencia a favor de la validez de constructo, coincidiendo la estructura obtenida con el planteamiento teórico. En primer lugar, en el Análisis factorial del instrumento completo los ítems se agruparon en seis factores, los cuales respetaron en gran medida la distribución de los distintos Componentes de la intención conductual. Sin embargo, los reactivos asociados a la Intención conductual y a la Conducta no formaron factores separados de dicha escala; razón por la cual se replicó el análisis, esta vez distinguiendo estas dimensiones a priori, puesto que de acuerdo a la Teoría de Conducta Planificada, estas escalas podrían considerarse instrumentos de medición individuales, con constructos separados, aludiendo a diferentes aspectos de la toma de decisiones volitiva (Ajzen, 2002; Eagly y Chaiken, 1993). El Análisis factorial por escala arrojó una estructura de seis factores para la escala de Componentes de la intención conductual, una estructura unifactorial para la escala Intención conductual y una de dos factores para la escala Conducta.

Al realizar un análisis específico de los factores de la escala de Componentes de la intención conductual, se observó que dos de ellos aludieron a la Actitud hacia la evaluación de aprendizajes. Estos factores se distinguieron conceptualmente entre sí por el enfoque de evaluación de aprendizajes al cual aludían: mientras uno de ellos se situó en el Modelo sistémico, relevando la importancia de la evaluación como proceso, el otro enfatizó la participación del alumnado dentro de éste a partir del Modelo comunicativo. Esta diferenciación coincide con los planteamientos de la literatura especializada (e.g., Ahumada, 2003; Deneen y Brown, 2016; Harris y Brown, 2013; Quinquer, 2000), en tanto la institucionalización de un nuevo enfoque evaluativo requiere de un cambio paradigmático de los diferentes actores implicados en el ámbito educativo. Particularmente, de acuerdo a Denem y Brown (2016), las concepciones de los docentes sobre la evaluación se fundamentan en sus conocimientos sobre la evaluación de aprendizajes y de sus experiencias emocionales con este proceso; ambos, aspectos definidos en las subescalas actitudinales del cuestionario desarrollado en este estudio.

Continuando con el análisis específico de los factores de la escala de Componentes de la intención conductual, otros dos factores dieron cuenta de la Norma subjetiva, es decir, la incidencia de las opiniones de terceros sobre las prácticas evaluativas del profesorado, considerando en un factor a los superiores y en el otro a los pares. Ello podría deberse a que la relación entre pares y la relación entre docente y jefatura es diferente en términos jerárquicos; en especial, cuando la concepción del profesorado acerca de la evaluación de aprendizajes, aún está ligada a una medición indirecta acerca de su propio desempeño (Brown et al., 2015); de este modo, resulta consistente analizar la disposición de los docentes a considerar las opiniones de sus colegas o de sus superiores, como factores separados.

Respecto de los factores que surgen a partir del análisis factorial de la subescala de Control percibido, dos factores aludieron a ello, diferenciándose entre sí de acuerdo al aspecto que influía en la concreción de determinadas prácticas evaluativas. De esta forma, un factor estuvo asociado al tiempo y otro a las competencias de los mismos docentes en el ámbito de la evaluación de aprendizajes. Considerando que la docencia constituye una de las profesiones más demandantes, en términos físicos, psicológicos, sociales y organizacionales (Fernet et al., 2012), especialmente en Chile dadas las condiciones laborales del profesorado (Cornejo, 2009; Cuenca et al., 2009; Ruffinelli, 2014), el factor tiempo constituye un aspecto relevante al momento de definir aspectos que pudiesen limitar la concreción de prácticas evaluativas de calidad. En cuanto a las competencias evaluativas, la literatura especializada da cuenta de la importancia de que los docentes cuenten con una formación robusta que les permita efectuar prácticas pedagógicas mayor calidad, de manera que los avances teóricos en esta materia puedan concretarse a nivel aula, traduciéndose así en una mejora en los aprendizajes de los estudiantes (DeLuca et al., 2013; DeLuca y Klinger, 2010; DeLuca et al., 2015; Deneen y Brown, 2016; McGee y Colby, 2014; Popham, 2011). En consecuencia, los aspectos limitantes hallados a partir del análisis factorial son coherentes con los planteamientos teóricos previamente desarrollados.

Tras efectuar los análisis psicométricos anteriores se propuso una versión final del cuestionario conformada por una selección de 30 ítems. Para ello se utilizaron primeramente los resultados del análisis factorial por escala, para luego escoger en cada factor aquellos ítems que no presentasen diferencias significativas entre grupos, de acuerdo al género o a la región, que aportasen consistencia interna al instrumento y cuya capacidad discriminativa fuese óptima. De este modo, la estructura factorial mantuvo la distribución dada por la teoría, al igual que en el Análisis factorial previo.

Por otra parte, más de un 75% de los ítems escogidos presentaron un índice de capacidad discriminativa sobre 0,30, considerándose un nivel adecuado dadas las características del cuestionario (Anastasi y Urbina, 1998; Cohen y Swerdik, 2000; Crocker y Algina, 1986); de este modo, los reactivos escogidos permitirían diferenciar entre los docentes que tienen una disposición positiva a efectuar prácticas evaluativas alineadas al Modelo comunicativo de evaluación de aprendizajes, de lo que presentan más reticencia a dicho enfoque.

Adicionalmente, ninguno de estos ítems presentó diferencias significativas entre grupos de acuerdo al género y a la región, con objeto de reducir la potencial existencia de sesgo producto de características demográficas que no guardan directa relación con el constructo que el instrumento intenta medir (Hogan, 2004; Martínez, 1995).

En consecuencia, se obtuvieron indicadores psicométricos satisfactorios para el cuestionario completo y para sus dimensiones, obteniéndose con ello evidencia a favor de su validez. La consistencia interna del cuestionario, la capacidad discriminativa de los reactivos que lo conforman, la coherencia de la estructura factorial y teórica, además de la ausencia de reactivos con diferencias significativas en variables de género y región, demuestran que la propuesta realizada constituye un instrumento robusto para la medición de intenciones conductuales del profesorado hacia la evaluación de aprendizajes.

VI. LIMITACIONES Y PROYECCIONES DEL CUESTIONARIO

6.1. Limitaciones del cuestionario

Una de las principales limitaciones de este cuestionario, referida a su construcción, se relacionó con el método de recolección de información escogido; particularmente el sesgo que puede llegar tener un instrumento de autorreporte. En este cuestionario, las respuestas de los participantes correspondieron a declaraciones explícitas, pudiendo estar sujetas a lo que los respondientes consideraban socialmente aceptable, sin que ello necesariamente se ajustase a sus disposiciones reales hacia la evaluación o a su comportamiento efectivo en el aula (Chan, 2009; Van de Mortel, 2008). Pese a estas falencias, la elección de un instrumento de autorreporte resultó apropiada en este caso, dado que el rasgo medido correspondía a un constructo referido a factores internos. En consecuencia, futuras investigaciones podrían robustecer la evidencia a favor de su validez, en términos de su relación con otras variables, contrastando la información con la observación de prácticas evaluativas en el aula o incorporando escalas de deseabilidad social que permitan controlar posibles sesgos.

En relación a la validación psicométrica de los reactivos, cabe señalar que una sus principales limitaciones consistió en el tipo de muestreo efectuado, en tanto este fue no probabilístico; en consecuencia, se realizaron Tests Chi-cuadrado para verificar la representatividad de la muestra respecto de la población, en términos de género de los docentes, años de experiencia y dependencia del establecimiento en el que se desempeñaban. En particular, el muestreo combinó la autoselección de sus participantes con un muestreo en cadena. De esta manera, considerando que la participación voluntaria podría generar un sesgo de selección, en tanto los encuestados podrían presentar una afinidad mayor con el tema de la evaluación de aprendizajes que el general de la población de estudio, se persiguió reducir este efecto, realizando un sorteo con incentivos económicos que permitiese promover la participación tanto de los docentes con interés hacia el tópico estudiado como sin interés hacia él. (Hernández, Fernández y Baptista, 2010).

Respecto del tamaño muestral, si bien se cumplió con la cantidad mínima para efectuar los análisis psicométricos aplicados en el presente estudio, es importante reconocer que éstos podrían robustecerse a futuro, contando con un tamaño muestral mayor que complementase estos resultados; por ejemplo, utilizando un modelamiento de puntuaciones bajo Teoría de Respuesta al Ítem, empleando dos muestras para contraponer los análisis factoriales exploratorio y confirmatorio o calculando estadísticos referidos al funcionamiento diferencial de los ítems. Pese a ello, los análisis presentados en este estudio constituyen evidencia contundente a favor del cuestionario construido.

6.2. Proyecciones del cuestionario

El instrumento desarrollado en el presente estudio constituirá una herramienta que proveerá información especialmente útil para la formación de futuros docentes, para el desempeño del profesorado en los centros educativos y para el desarrollo de políticas públicas en educación. En relación a ello, este cuestionario tiene como principal propósito dar cuenta de la estructura de pensamiento interna del profesorado, distinguiendo las diferentes motivaciones que finalmente lo llevan a concretar determinadas prácticas evaluativas en el aula. De este modo, la información que proveerá dicho instrumento podrá dirigir de mejor manera los recursos aludiendo a iniciativas focalizadas de formación continua y acompañamiento docente, en función de la mejora de las competencias en evaluación de aprendizajes del profesorado.

Otro uso práctico asociado al instrumento desarrollado en este estudio, corresponde al aporte que éste tiene para la investigación en materia evaluativa. A partir de la construcción y validación efectuada por medio de este estudio, es posible contar con un cuestionario en español validado en Chile, que puede ser empleado en futuras investigaciones. En este marco, resulta relevante destacar su potencial uso en la detección de cambios en las disposiciones de los docentes hacia la evaluación de aprendizajes, aplicando el cuestionario antes y después de alguna intervención dirigida a la mejora de sus competencias evaluativas, detectando el efecto de ésta en su estructura interna de pensamiento. Asimismo, el instrumento también puede aplicarse correlacionando sus resultados con los relativos al conocimiento que tiene el profesorado sobre la evaluación de aprendizajes, a modo de contrastar hipótesis que relacionan las motivaciones intrínsecas y la alfabetización de los docentes en este tópico.

Por otra parte, se podría estudiar más profundamente el poder predictivo de las intenciones conductuales sobre la ejecución de determinadas prácticas evaluativas, distinguiendo qué componentes resultan más críticos en función de su mejora.

Adicionalmente, considerando la dificultad que tiene la instauración de un modelo de evaluación de aprendizajes en la práctica, resulta también interesante medir por separado, en futuras investigaciones, la intención conductual desde el modelo sistémico, en tanto este modelo propone una visión de la evaluación como un proceso orientado a la mejora de los aprendizajes; y por otro, la intención conductual integrando al alumnado en el proceso evaluativo, es decir, desde el modelo comunicativo. De este modo, podría recabarse información más precisa desde distintas perspectivas acerca de la evaluación, transitando hacia una concepción en la que los estudiantes participan más activamente.

REFERENCIAS BIBLIOGRÁFICAS

- Ahumada, P. (2002). *La evaluación en una concepción de aprendizaje significativo*. Valparaíso, Chile: Ediciones Universitarias de Valparaíso de la Universidad Católica de Valparaíso.
- Ahumada, P. (2003). El estilo evaluativo y su incidencia en las prácticas de docentes universitarios. *Pensamiento Educativo*, 32(1), 264-285.
- Ahumada, P. (2005). La evaluación auténtica: un sistema para la obtención de evidencias y vivencias de los aprendizajes. *Revista perspectiva educacional*, (45), 11-24.
- Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment: Concepts and applications* (6th ed.). New York, U.S.A.: McGraw Hill.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.
- Ajzen, I. (2002). Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology*, 32(4), 665-683.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. New Jersey, U.S.A.: Prentice-Hall.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. En D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). New Jersey, U.S.A.: Erlbaum.
- Allport, G. W. (1935). Attitudes. En C. Murchison (Ed.), *A handbook of social psychology* (pp. 798-844). Massachusetts, U.S.A.: Clark University Press.

- Allport, G. W. (1985). The historical background of social psychology. En G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., Vol. 1, pp. 1-46). New York, U.S.A.: McGraw Hill.
- Anastasi, A., y Urbina, S. (1998). *Test psicológicos*. Ciudad de México, México: Prince Hall.
- Anijovich, R. (2013). La retroalimentación en la evaluación. En R. Anijovich (Comp.), *La evaluación significativa* (pp. 129-149). Buenos Aires, Argentina: Paidós.
- Aznar, P. (1992). El constructivismo en la Educación. En P. Aznar (Coord.), *Constructivismo y Educación* (pp. 13-32). Valencia, España: Tirant lo blanch.
- Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2015). The relation between Assessment for Learning and elementary students' cognitive and metacognitive strategy use. *British Journal of Educational Psychology*, 85(1), 33-46.
- Bagozzi R. P., & Burnkrant R. E., (1979), Attitude Measurement and Behaviour Change: A Reconsideration of Attitude Organization and Its Relationship to Behaviour, *Advances in Consumer Research*, 6, 295–302.
- Baird, J. (2010). Beliefs and practice in teacher assessment. *Assessment in Education: Principles, Policy & Practice*, 17(1), 1-5.
- Barkley, E. F., & Major, C. H. (2015). *Learning Assessment Techniques: A Handbook for College Faculty*. California, U.S.A.: John Wiley & Sons.
- Beck, J., Czerniak, C. M., & Lumpe, A. T. (2000). An exploratory study of teachers' beliefs regarding the implementation of constructivism in their classrooms. *Journal of Science Teacher Education*, 11(4), 323-343.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Berkshire, U.K.: McGrawHill Education.

- Black, P., Harrison, C., & Lee, C. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9–21.
- Brahm, T., & Euler, D. (2013). *Development of Students' Attitudes during their First Year at a Business School*. Recuperado de https://www.alexandria.unisg.ch/223370/1/1305_brahm_euler_efmd-he-research.pdf
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11(3), 301-318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports* 99(1), 166-170.
- Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3, 45-70.
- Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research*, 71, 50-64.
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 347-363.
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27(1), 210-220.

- Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education*, 26(3), 319-337.
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish journal of psychology*, 15(1), 75-89.
- Carretero, M. (2005). *Constructivismo y educación*. Ciudad de México, México: Progreso.
- Casanova, M. A. (1998). *La evaluación educativa*. Ciudad de México, México: SEP-Muralla.
- Cattell, R. (1966). The Scree Test for the number of factors. *Multivariate Behavioral Research*. 1(2), 141-161.
- Chan, D. (2009). So why ask me? Are self-report data really that bad. En C. E. Lance, & R. J. Vandenberg (Eds.) *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309-336). New York, U.S.A.: Taylor & Francis.
- Cohen, R., y Swerdlik, M (2000). *Pruebas y evaluación psicológicas* (4ª ed.). Ciudad de México, México: McGraw Hill.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2ª ed.). New Jersey, U.S.A.: Lawrence Erlbaum Associates.
- Conatser, P., Block, M., & Gansneder, B. (2002). Aquatic instructors' beliefs toward inclusion: the theory of planned behavior. *Adapted Physical Activity Quarterly*, 19(2), 172-187.

- Cornejo, R. (2009). Condiciones de trabajo y bienestar/malestar docente en profesores de enseñanza media de Santiago de Chile. *Educação & Sociedade*, 30(107), 409-426.
- Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from your Analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1-9.
- CPEIP y MINEDUC. (2008). *Marco para la buena enseñanza*. Santiago, Chile: Autor. Recuperado de <http://www.docentemas.cl/docs/MBE2008.pdf>
- Crawley, F. E. (1990). Intentions of science teachers to use investigative teaching methods: A test of the theory of planned behavior. *Journal of Research in Science Teaching*, 27(7), 685-697.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida, U.S.A.: Harcourt Brace Jovanovich.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cuenca, R., Fabara, E., Kohen, J., Parra, M., Rodríguez, L. y Tomasina, F. (2009). *Condiciones de trabajo y salud docente: estudios de casos en Argentina, Chile, Ecuador, México, Perú y Uruguay*. Santiago, Chile: UNESCO. Recuperado de <http://www.bvcooperacion.pe/biblioteca/bitstream/123456789/4774/1/BVCI0003968.pdf>
- Danielson, C. (2008). Assessment for learning: For teachers as well as students. En C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 191-213). New York, U.S.A.: Taylor & Francis.
- Danielson, C. (2013). *Framework for teaching evaluation instrument*. Princeton, New Jersey, U.S.A.: The Danielson Group. Recuperado de

<http://usny.nysed.gov/rttt/teachers-leaders/practicerubrics/Docs/danielson-teacher-rubric.pdf>

- De la Torre, S. (1998). Estilos de evaluación: ¿Cómo explicar las diferencias entre los profesores? En A. Medina, J. Cardona, S. Castillo y M. C. Domínguez (Eds.), *Evaluación de los procesos y resultados del aprendizaje de los estudiantes* (pp. 333-345). Madrid, España: UNED.
- De la Torre, S. (2002). Aprender de los errores: el tratamiento didáctico de los errores como estrategia de innovación. En S. Castillo (Coord.), *Compromisos de la evaluación educativa*. Madrid, España: Prentice Hall.
- DeLuca, C. (2012). Preparing Teachers for the Age of Accountability: Toward a Framework for Assessment Education. *Action in Teacher Education*, 34(5-6), 576-591.
- DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for Preservice Assessment Education: Supporting Teacher Candidates' Assessment Literacy Development. *The Teacher Educator*, 48(2), 128-142.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2015). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 1-22.
- Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, 3(1), 1225380.

- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13(3), 440-464.
- Dochy, F. (2009). The Edumetric Quality of New Modes of Assessment: Some Issues and Prospects. En G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 85-114). Wollongong, Australia: Springer.
- Eagly A. H., & Chaiken S. (1993). *The psychology of attitudes*. Florida, U.S.A.: Harcourt Brace Jovanovich College Publishers.
- Elige Educar. (2015). Radiografiara de los profesores de las escuelas en Chile [Infografía]. Recuperado de http://eligeeducar.cl/wp-content/uploads/2015/10/radiografia_profesores.jpg
- Everitt, B. S. y Wykes, T. (2001). *Diccionario de Estadística para Psicólogos*. Madrid, España: Ariel.
- Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education*, 28(4), 514-525.
- Fiske, S. T. (2010). *Social beings: Core motives in social psychology* (2da ed.). New Jersey, U.S.A.: John Wiley & Sons.
- Förster, C. y Rojas-Barahona, C. (2008). Evaluación al interior del aula: una mirada desde la validez, confiabilidad y objetividad. *Revista Pensamiento Educativo* 43(2008), 285-305.
- Franzoi, S. (2010). *Psychology: A discovery experience*. Ohio, U.S.A.: Cengage Learning.

- Gebril, A., & Brown, G. T. L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy and Practice*, 21(1), 16-33.
- Glasersfeld, E. von (1989). *Cognition, Construction of Knowledge, and Teaching. Synthese*, 80(1), 121-140.
- Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, 138(6), 1085.
- Gorsuch, R. (1983). *Factor analysis*. New Jersey, U.S.A: Erlbaum
- Greenwood, J. D. (2015). *A Conceptual History of Psychology* (2nd ed.). Cambridge, U.K.: Cambridge University Press.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. California, U.S.A.: SAGE.
- Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. (1999). *Análisis Multivariante*. Madrid, España: Prentice Hall.
- Haladyna, T. (1999). *Developing and Validating multiple-choice test items*. New Jersey, U.S.A.: Lawrence Erlbaum.
- Harasim, L. (2012). *Learning Theory and Online Technologies*. Missouri, U.S.A.: Walsworth Publishing Company.
- Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer-and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111.
- Hernández, R., Fernández, C. y Baptista, P. (2010). *Metodología de la investigación*. Ciudad de México, México: McGraw Hill.

- Herrada, M. y Zapata, A. (2015). ¿Qué nos dice el portafolio acerca de las mejoras en el desempeño docente? *Midevidencias 1*, 1-7. Recuperado de:
<http://www.mideuc.cl/wp-content/uploads/2015/06/MidEvidencias-N1.pdf>
- Hess, J., & Singer, E. (1995). The role of respondent debriefing questions in questionnaire development. En American Statistical Association, *Proceedings of the Section on Survey Research Methods, American Statistical Association* (Survey Research Methods Section). Virginia, U.S.A: Autor.
- Hoffmann, J. (2013). Evaluación mediadora: una propuesta fundamentada. En R. Anijovich (Comp.), *La evaluación significativa* (pp. 73-102). Buenos Aires, Argentina: Paidós.
- Hogan, T. (2004). *Pruebas psicológicas: una introducción práctica*. Ciudad de México, México: Manual Moderno.
- Horn, J. (1965). A rationale and test for the number the factors in factor analysis. *Psychometrika*, 30(2), 179-185.
- Hughes, K. A. (2004). Comparing pretesting methods: Cognitive interviews, respondent debriefing, and behavior coding. *Survey Methodology*, 2, 1-20.
- IBM Corporation (2012). *IBM SPSS Statistics* (Version 21.0). New York, U.S.A.: Author.
- Jorba, J. y Sanmartí, N. (2000). La función pedagógica de la evaluación. En M. Rovira (Comp.), *Evaluación como ayuda al aprendizaje* (pp. 13-20). Barcelona, España: Graó.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

- Kim, K. H., Lu, J., & Estrada-Hernandez, N. (2015). Attitudes Toward People With Disabilities: The Tripartite Model, Social Desirability, and Other Controversial Variables. *Journal of Asia Pacific Counseling*, 5(1), 23-37.
- Kline, P. (2000). *Handbook of Psychological Testing*. London, U.K.: Routledge.
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research, and Evaluation*, 12(2), 1-11.
- Lee, J., Cerreto, F. A., & Lee, J. (2010). Theory of Planned Behavior and Teachers' Decisions Regarding Use of Educational Technology. *Educational Technology & Society*, 13 (1), 152-164.
- Ley N° 18.962 (1990). *Ley Orgánica Constitucional de Enseñanza*. Diario oficial de la República de Chile. Recuperado de <http://www.leychile.cl/Navegar?idNorma=30330>
- Ley N° 20.370. (2009). *Establece la Ley General de Educación*. Diario oficial de la República de Chile. Recuperado de <http://www.leychile.cl/Navegar?idNorma=1006043>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of personality assessment*, 94(2), 196-204.
- Lorenzo-Seva, U., & Ferrando, P. J. (2015). *FACTOR* (Versión 10.3.01). Recuperado de <http://psico.fcep.urv.es/utilitats/factor/soft/factor10.3.64bits.zip>

- Mann, H. B., Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18 (1): 50-60.
- Mardia, K. V. (1970), Measures of multivariate skewnees and kurtosis with applications. *Biometrika*, 57, 519-530.
- Martin, J. J., & Kulinna, P. H. (2004). Self-efficacy theory and the theory of planned behavior: Teaching physically active physical education classes. *Research quarterly for exercise and sport*, 75(3), 288-297.
- Martínez, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid, España: Síntesis.
- McCoy, J. (2014). *Change Model 83 Success Secrets - 83 Most Asked Questions On Change Model - What You Need To Know*. Queensland, Australia: Emero Publishing. Recuperado de <https://books.google.cl/books?id=1SINBwAAQBAJ&printsec=frontcover&hl=es#v=onepage&q&f=false>
- McGee, J., & Colby, S. (2014). Impact of an Assessment Course on Teacher Candidates' Assessment Literacy. *Action in Teacher Education*, 36(5-6), 522-532.
- McPhail, G. (2015). The fault lines of recontextualisation: The limits of constructivism in education. *British Educational Research Journal*, 42 (2), 294-313
- Mergel, B. (1998). *Instructional Design and Learning Theory*. Recuperado de <http://etad.usask.ca/802papers/mergel/mergel.pdf>
- Meyer, J. P. (2014). *JMetrik* (Version 3.1.2). Recuperado de http://www.itemanalysis.com/jmetrik-3_1_2/jmetrik_windows_3_1_2.exe
- Mills, J. A. (2000). *Control: A History of Behavioral Psychology*. New York, U.S.A.: New York University Press.

- MINEDUC. (2008). *Planificación y Diseño de la Enseñanza*. Santiago, Chile: Autor.
Recuperado de <http://www.e-historia.cl/biblioteca/Mineduc/Planificaci%C3%B3n%20y%20Dise%C3%B1o%20de%20la%20Ense%C3%B1anza.pdf>
- MINEDUC. (2009a). *Evaluación Para el Aprendizaje: Educación Básica Primer Ciclo*. Santiago, Chile: Autor. Recuperado de <http://www.cpeip.cl/usuarios/cpeip/File/2013materialespsp/1cicloepa.pdf>
- MINEDUC. (2009b). *Evaluación Para el Aprendizaje: Educación Básica Segundo Ciclo*. Santiago, Chile: Autor. Recuperado de <http://www.cpeip.cl/usuarios/cpeip/File/2013materialespsp/2cicloepa.pdf>
- MINEDUC. (2011a). *Bases Curriculares. Educación Básica: Lenguaje y Comunicación, Matemática, Ciencias Naturales, Historia, Geografía y Ciencias Sociales, Idioma Extranjero Inglés*. Recuperado de http://www.curriculumenlineamineduc.cl/605/articles-22394_programa.pdf
- MINEDUC. (2011b). *Fundamentos. Bases Curriculares 2011. Educación Básica. Propuesta presentada para aprobación del Consejo Nacional de Educación*. Recuperado de http://files.aprotec-umce7.webnode.cl/200000765-7f3d981323/Fundamentaci%C3%B3n%20Bases%20Curriculares%20Ens_Bas_MINEDUC_Ago_2011.pdf
- MINEDUC. (2011c). *Resultados Evaluación Docente 2010* [Diapositiva]. Recuperado de http://www.docentemas.cl/docs/2011/Resultados_Ev_Docente_2010_29032011.pdf
- MINEDUC. (2012). *Orientaciones para evaluar los aprendizajes*. Recuperado de <http://www.curriculumenlineamineduc.cl/605/w3-article-14602.html>

- MINEDUC. (2014). *Evaluación del Desempeño Profesional Docente. Resultados 2013* [Diapositiva]. Recuperado de http://www.docentemas.cl/docs/Resultados_Evaluacion_Docente_2013.pdf
- MINEDUC. (2015). *Resultados Evaluación Docente 2014* [Diapositiva]. Recuperado de http://www.docentemas.cl/docs/Resultados_Evaluacion_Docente_2014.pdf
- Miranda, A. (1998). *Epistemología de la psicología*. Ciudad de México, México: UNAM.
- Muthén, B. y Kaplán, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189.
- Muthén, B., & Kaplan D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45(1), 19-30.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6ta ed.). New Jersey, U.S.A.: Pearson/Prentice Hall.
- OECD. (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. Paris, Francia: OECD Publishing.
- Ordoñez, C. L. (2004). Pensar pedagógicamente desde el constructivismo. De las concepciones a las prácticas pedagógicas. *Revista de estudios sociales*, (19), 7-12.
- Parra, E. (2008). *Evaluación para los aprendizajes y la enseñanza*. Santiago, Chile: Ediciones Universidad Cardenal Silva Henríquez.
- Pérez, E. R., & Medrano, L. (2010). Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina de Ciencias del Comportamiento* 2(1), 58-66.

- Pérez, A., Soto, E., Sola, M. y Serván, M. J. (2009). *Orientar el desarrollo de competencias y enseñar cómo aprender* (Vol. 4). Madrid, España: AKAL.
- Pérez Juste, R. (2006). *Evaluación de Programas Educativos*. Madrid, España: La Muralla.
- Perrenoud, P. (2008). *La evaluación de los alumnos: De la producción de la excelencia a la regulación de los aprendizajes. Entre dos lógicas*. Buenos Aires, Argentina: Colihue.
- Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational researcher*, 24(7), 5-12.
- Piaget, J. (1970). Piaget's Theory. En P. H. Mussen (Ed.). *Carmichael's Manual of Child Psychology* (Vol. 1, pp. 703-732). New York, U.S.A.: Wiley.
- Picado, F. M. (2006). *Didáctica General: una perspectiva integradora*. San José, Costa Rica: EUMED.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, 30(4), 459-467.
- Popham, W. J. (2011). Assessment Literacy Overlooked: A Teacher Educator's Confession. *The Teacher Educator*, 46(4), 265-273.
- Quinquer, D. (2000). Modelos y enfoques sobre la evaluación: el modelo comunicativo. En M. Rovira (Comp.), *Evaluación como ayuda al aprendizaje* (pp. 13-20). Barcelona, España: Graó.
- Reyes, L. (2007). La Teoría de la Acción Razonada. Implicaciones para el estudio de las actitudes. *Investigación Educativa Duranguense*, (7), 66-77.

- Ritchey, F. J. (2002). *Estadística para las ciencias sociales: El potencial de la imaginación estadística*. Ciudad de México, México: McGraw-Hill/Interamericana.
- Rivas-Ruiz, R., Moreno-Palacios, J., & Talaveraa, J. O. (2013). Diferencias de medianas con la U de Mann-Whitney. *Rev Med Inst Mex Seguro Soc*, 51(4), 414-419.
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. En M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). New Haven, Connecticut, United States: Yale University Press.
- Ruffinelli, A. (2014). Los imprescindibles para una nueva profesión docente en Chile. *Cuaderno de Educación*, 61.
- Sánchez, J. G. y Martínez, L. (1998). Viejas teorías - Teorías actuales del aprendizaje. En H. S. Torres y A. Miranda (Comps.), *Crisis y emergencia de paradigmas en psicología* (pp. 42-50). Ciudad de México, México: UNAM.
- Sanmartí, N. (2007). *10 Ideas clave: Evaluar para aprender*. Barcelona, España: Graó.
- Santiago, P. (2013, noviembre). *Evaluación docente en Chile: Fortalezas y desafíos. WinnRevisión de la OCDE de los marcos de evaluación para mejorar los resultados escolares* [Diapositiva]. Ponencia presentada en el Seminario Evaluación Docente en Chile: Un enfoque internacional, Santiago, Chile. Recuperado de http://www.docentemas.cl/docs/Presentacion_Revision_Ev_Docente_Chile_Paul_o_Santiago_Recomendaciones2_08112013.pdf
- Santiago, P., Benavides, F., Danielson, C., Goe, L. & Nusche, D. (2013). *OECD Reviews of Evaluation and Assessment in Education Teacher Evaluation in Chile 2013*. Paris, France: OECD Publishing. Recuperado de

http://www.docentemas.cl/docs/Teacher_Evaluation_in_Chile_OECD_Santiago_et_al_Nov2013.pdf

Santibáñez, J. D. (2002). Prólogo. En P. Ahumada, *La evaluación en una concepción de aprendizaje significativo* (pp. 7-10). Valparaíso, Chile: Ediciones Universitarias de Valparaíso de la Universidad Católica de Valparaíso.

Santos Guerra, M. A. (2005). Evaluar es comprender: de la dimensión técnica a la dimensión crítica. *Revista Investigaciones en Educación* 5(1), 67-85.

Seattler, P. (2004). *The evolution of American educational technology*. Colorado, U.S.A.: Libraries Unlimited.

Soler, E. (2006). *Constructivismo, innovación y enseñanza efectiva*. Caracas, Venezuela: Equinoccio.

Stanovich, P. J., & Jordan, A. (1998). Canadian teachers' and principals' beliefs about inclusive education as predictors of effective teaching in heterogeneous classrooms. *The Elementary School Journal*, 221-238.

Stiggins, R. J., & Chappuis, J. (2012). *An introduction to student-involved assessment for learning*. Boston, United States: Pearson.

Schwartz, R. M., Schmitt, M. C., & Lose, M. K. (2012). Effects of teacher-student ratio in response to intervention approaches. *The Elementary School Journal*, 112(4), 547-567.

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics*. New York, U.S.A.: Harper & Row.

Taber, C. S., & Young, E. (2011). Information processing, public opinion, and accountability. En S. Odugbemi y T. Lee (Eds.), *Accountability through Public Opinion: From Inertia to Public Action* (pp. 95-122). Washington D.C., U.S.A.: World Bank Publications.

- Tejada, J. (1999). La evaluación: su conceptualización. En B. Jiménez (Ed.), *Evaluación de programas, centros y profesores* (pp. 25-56). Madrid, España: Síntesis.
- Teo, T., & Lee, C. B. (2010). Explaining the intention to use technology among student teachers: An application of the Theory of Planned Behavior (TPB). *Campus-Wide Information Systems*, 27(2), 60-67.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods*, 16(2), 209-220.
- Theodorakis, Y., Bagiatas, K., & Goudas, M. (1995). Attitudes toward teaching individuals with disabilities: Application of planned behavior theory. *Adapted Physical Activity Quarterly*, 12, 151-160.
- Valverde, G. (2000). La interpretación justificada y el uso apropiado de los resultados de las mediciones. En P. Ravela (Ed.), *Los Próximos Pasos: ¿Hacia Dónde y Cómo Avanzar en la Evaluación de Aprendizajes en América Latina?* (pp. 21-30). Lima, Perú: GRADE/Preal.
- Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, 25(4), 40-48.
- Vardar, E. (2010). *Sixth, seventh and eighth grade Teachers' Conception of Assessment* (Doctoral dissertation). Middle East Technical University, Ankara, Turkey.
- Vygotsky, L. (1978). *Mind in society*. Massachusetts, U.S.A.: Harvard University Press.
- William, D. (2011). *Embedded formative assessment*. Indiana, U.S.A.: Solution Tree Press.
- Winn, W. (2003). Cognitive perspectives in psychology. En D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology: A*

Project of the Association for Educational Communications and Technology
(2nd ed., pp. 79-112). New York, U.S.A.: Simon & Schuster Macmillan.

Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education*, 45, 128-136.

Zabalza, M. A. (1998). Evaluación de actitudes y valores. En A. Medina, J. Cardona, S. Castillo y M. C. Domínguez (Eds.), *Evaluación de los procesos y resultados del aprendizaje de los estudiantes* (pp. 245-300). Madrid, España: UNED.

Zabalza, J. (2007). *Las estrategias evaluativas en el proceso educativo*. Recuperado de <http://ww2.educarchile.cl/UserFiles/P0001/File/Estrategias%20evluativas.pdf>