

DISEÑO Y VALIDACIÓN DE UN TEST PARA LA EVALUACIÓN DE COMPETENCIAS EN LA FORMACIÓN INICIAL DEL PROFESORADO DE CIENCIAS Y MATEMÁTICA

DESIGN AND VALIDATION OF A TEST FOR THE EVALUATION OF COMPETENCES IN THE INITIAL TRAINING OF THE TEACHER OF SCIENCES AND MATHEMATIC

Yonnhatan García C., Juan Vargas M., Guillermo Arancibia C., Claudio Pérez M., Isabel Vargas C., Agustín Adúriz B..

Resumen

En este artículo se describe el proceso para construir y validar un test que evalúa algunas competencias científicas y matemáticas en estudiantes que ingresan a carreras de Pedagogía en Ciencias (Física, Química, Biología) y Pedagogía en Matemática para la educación secundaria. Los criterios utilizados para establecer la idoneidad psicométrica de los ítems se basaron en la teoría clásica de los tests. En general, el test muestra una dificultad media ($p = 54\%$). El índice de discriminación, estimado a partir de una correlación biserial, indica que el test discrimina bien el rasgo evaluado ($r_b > 3,5$). También se realizó un análisis de confiabilidad mediante el estadístico alpha de Cronbach ($\alpha > 0,9$), que indica alta confiabilidad del test final compuesto por los ítems que superaron el pilotaje.

Palabras clave: evaluación de competencias; test de competencias; formación de profesores; didáctica de las ciencias; didáctica de la matemática.

Abstract

In this article we describe the process to build and validate a test that evaluates some scientific and mathematical competences in students who enter careers in Pedagogy in Sciences (Physics, Chemistry, Biology) and Pedagogy in Mathematics for secondary education. The criteria used to establish the psychometric suitability of the items were based on the classical theory of the tests. In general, the test shows a medium difficulty ($p = 54\%$). The discrimination index, estimated from a biserial correlation, indicates that the test discriminates well the evaluated trait ($r_b > 3.5$). A reliability analysis was also carried out using the Cronbach alpha statistic ($\alpha > 0.9$), which indicates high reliability of the final test composed of the items that exceeded piloting.

Keywords : evaluation of competences; competency test; teacher training; didactics of science; didactics of mathematics.

Antecedentes

En el marco de un proyecto¹ desarrollado entre los años 2014 y 2016, se diseñó un modelo para la nivelación de competencias científicas basado en la indagación. Dicho modelo fue implementado en estudiantes que ingresaban a la Formación Inicial Docente (FID) en carreras de Pedagogía en Física, Química y Biología, y Pedagogía en Matemática de la Universidad Metropolitana de Ciencias de la Educación (UMCE, Santiago, Chile). El proyecto implicaba la nivelación de tres competencias generales en el contexto de las ciencias y matemáticas: Modelamiento-Modelización, Razonamiento Lógico, y Argumentación. Dada la necesidad de nivelar, una cuestión fundamental era determinar el nivel de estas tres competencias cuando los estudiantes ingresan a la FID. En la misma línea, en un estudio realizado sobre la misma población (Reyes & García-Cartagena, 2014) se logró observar que una de las componentes de las competencias científicas evaluadas se manifestaron de forma transversal y sin diferencias significativas en relación con la carrera de origen de los estudiantes. Finalmente, disponer de esta información permite conocer de mejor manera el perfil de ingreso de los estudiantes, para con ello poder hacer los ajustes curriculares correspondientes.

Dimensiones de evaluación del test: competencias científicas

Largo ha sido el debate en torno a qué entender por competencia o por "ser competente". Según la Red Europea de Información sobre la Educación (Eurydice, 2002), una competencia es definida como la capacidad de realizar una tarea en un contexto determinado, con lo cual se incorpora la idea de que un individuo logrará el desarrollo de las competencias siempre y cuando se enfrente a situaciones que le permitan aplicar esquemas aprendidos en un nuevo contexto².

Por otra parte, la OCDE, a partir del proyecto DeSeCo (OCDE, 2002), se encargó de definir y seleccionar las competencias consideradas esenciales para la vida de las personas y el buen funcionamiento de la sociedad. Para la OCDE (2002), las competencias son una combinación de habilidades prácticas, conocimientos, motivación, valores éticos, actitudes, emociones y otros componentes sociales y de comportamiento que

se movilizan conjuntamente para lograr una acción eficaz. Estas competencias han servido de base para la elaboración de evaluaciones internacionales como PISA y para la incorporación de políticas educacionales que pretenden mejorar el logro de las competencias definidas. Asimismo, se suele enfatizar que las competencias son clave cuando resultan valiosas para la población, independiente del género, la condición socio-cultural y el entorno familiar (OCDE, 2002).

Por su parte, el proyecto Tuning Latinoamérica (Beneitone et al., 2007) propone definir el concepto de competencia como la capacidad del individuo para: i) conocer y comprender, lo que involucra el conocimiento teórico; ii) saber cómo actuar, lo que considera una aplicación práctica y operativa en ciertas situaciones; y, finalmente, iii) saber cómo ser, lo que involucra el percibir a otros y vivir en un contexto social. Se sostiene entonces que las competencias, así como la inteligencia, no son una capacidad innata, sino que, por el contrario, son susceptibles de ser desarrolladas y construidas a partir de las motivaciones internas de cada cual, motivaciones que deberán ser comunicadas al grupo de trabajo (Beneitone et al., 2007).

En cuanto a las competencias científicas, la OCDE (2006), en el marco de la prueba PISA, sugiere que ellas involucran el conocimiento científico y el uso que se hace de ese conocimiento para identificar cuestiones, adquirir nuevos conocimientos, explicar fenómenos científicos y extraer conclusiones basadas en pruebas. En esta misma línea, Monereo & Pozo (2007) proponen que una competencia científica es entendida como la capacidad de reorganizar lo aprendido para transferirlo a nuevas situaciones y contextos. Desde esta perspectiva, una cuestión fundamental al promover el desarrollo de competencias científicas es que estas permitan a los sujetos tomar decisiones contextualizadas (García-Cartagena & Reyes, 2012). En este trabajo, como ha establecido, se plantea que las competencias científicas involucran al menos las capacidades, habilidades, destrezas y actitudes necesarias para la aplicación de cierto conocimiento científico de forma efectiva en contextos diversos.

En el contexto chileno, el Ministerio de Educación (MINEDUC, 2009) define competencias como sistemas de acción complejos que interrelacionan habilidades prácticas y cognitivas, conocimientos, motivación, orientaciones valóricas, actitudes y emociones, que en conjunto se movilizan para realizar una acción efectiva. Si bien el currículum nacional entrega esta definición de competencia, no declara abiertamente las competencias esperables en los estudiantes que egresan del sistema escolar chileno. La única aproximación a ellas se establece mediante la definición de los Objetivos Fundamentales que contemplan las habilidades a desarrollar, y de estas se puede llegar a definir implícitamente

1 Proyecto MECESUP UMC1302. "Plan de Mejoramiento para la nivelación de competencias científicas aplicado a estudiantes de primer año de Formación Inicial Docente de la Facultad de Ciencias Básicas basado en el modelo de Indagación Científica (Scientific Inquiry)". UMCE-MINEDUC, Chile (UMCE, 2013)

2 Aparecen entonces, como elementos constituyentes de toda competencia robusta, las tres "C": *capacidad* (cognitiva, discursiva, material, afectiva) de hacer algo sobre un *contenido* (científico, matemático) determinado, dentro de un *contexto* delimitado y reconocible (Adúriz-Bravo, 2012).

un “perfil de competencias” para el término de la enseñanza secundaria, con el cual el estudiantado se enfrentará a la Educación Superior.

El test que se presenta en este trabajo fue diseñado para medir el nivel de logro de tres competencias “científico-matemáticas” en estudiantes que ingresan a la FID en ciencias y matemática. Cada competencia evaluada fue obtenida a partir del cruce entre el ajuste curricular para la educación secundaria en Chile (MINEDUC, 2009), los

estándares pedagógicos y disciplinarios para la formación de docentes de secundaria (MINEDUC, 2012), el perfil de egreso universitario de los estudiantes de las carreras involucradas en el proyecto, y un debate teórico con especialistas a nivel nacional e internacional. A partir de este trabajo, se obtuvieron los descriptores de cada competencia, y sus estándares, niveles e indicadores (ver tabla 1), todos ellos insumos fundamentales para determinar qué medir en la construcción del test.

Tabla 1: Competencias científico-matemáticas consideradas para su nivelación.

Competencia	Descriptor	Estándar	Niveles	Indicadores
1. Razonamiento lógico	Procesamiento de representaciones mentales, datos e informaciones para llegar a propuestas o conclusiones basadas en procesos de razonamiento lógico utilizados en la construcción de conocimiento científico-matemático	Representa diversos procesos de pensamiento lógico que “desarrolla”/utiliza durante/ para la búsqueda y descubrimiento de relaciones para construir posibles predicciones, hipótesis o explicaciones a fenómenos del dominio de las ciencias experimentales y de la matemática	Nivel 1: Moviliza la información y conocimientos para la construcción de conocimientos complejos y/o la resolución de situaciones problema	Observa, selecciona
			Nivel 2: Evalúa problemas y controversias demostrando la evidencia disponible y formulando conjeturas	Organiza, clasifica
			Nivel 3: Analiza la información e identifica modelos	Analiza la información, representa
2. Modelización	Habilidad intelectual característica del pensamiento científico que se manifiesta en la elaboración de representaciones mentales o conceptuales que intentan explicar una situación o fenómeno observado estableciendo relaciones entre variables y parámetros	Construye representaciones para simular, describir y explicar las relaciones de los conceptos y procesos implicados en la construcción de proposiciones, modelos y teorías de las ciencias experimentales y de la matemática	Nivel 1: Propone explicaciones tentativas o preliminares en base a la organización e interpretación de datos estableciendo relaciones entre variables y/o parámetros	Formula conjeturas o hipótesis, y/o modelos
			Nivel 2: moviliza el conocimiento relacionando y aplicando conceptos y estrategias en la resolución de problemas	Procesa datos, diseña e implementa estrategias, contrasta conjeturas o hipótesis
			Nivel 3: resuelve problemas mediante la utilización y/o creación de modelos	Transfiere el modelo a situaciones novedosas

Competencia	Descriptor	Estándar	Niveles	Indicadores
3. Argumentación	Capacidad de construir relaciones sustantivas entre modelos y evidencias, relacionar datos y conclusiones, evaluar enunciados teóricos a la luz de los datos empíricos o procedentes de otras fuentes. Se basa en mecanismos de naturaleza cognitiva, metacognitiva, afectiva y social utilizados para justificar o refutar una opinión y/o hacer declaraciones teniendo en cuenta al receptor y la finalidad con la cual se emiten	Desarrolla explicaciones acerca de la interacción entre las ideas teóricas y la evidencia que las apoya, a partir de procesos de pensamiento lógico y sobre la base de valores compartidos, para justificar posturas acerca de ideas científico-matemáticas en discusión y/o la propia reconstrucción de los conocimientos adquiridos	Nivel 1: relaciona modelos y datos mediante la evaluación de evidencia e identificación de patrones	Relaciona datos, modelos y evidencia
			Nivel 2: evalúa problemas y controversias demostrando veracidad de la evidencia disponible y formulando conjeturas	Evalúa información y formula conjeturas
			Nivel 3: justifica conjeturas fundamentales como solución de un problema, ejemplifica dando cuenta de la relación teoría-evidencia para sostener la fundamentación sobre la solución de un problema	Justifica, comunica

Fundamento del diseño del test

El diseño general se realizó considerando la teoría clásica del test (TCT), y siguiendo las fases generales del proceso de construcción de instrumentos de medida (Carretero-Dios & Pérez, 2005; Downing & Haladyna, 2006; Muñiz & Fonseca-Pedrero, 2008). En la fase de pilotaje, el test contempló dos formas con 60 preguntas cada una. Las preguntas fueron de selección múltiple con 4 opciones de respuesta, de las cuales solo una de ellas sería la clave. El estímulo se realiza mediante la contextualización de situaciones problemáticas consideradas en el currículo nacional para los subsectores de Biología, Física, Química y Matemática (MINEDUC, 2009). En síntesis, cada ítem evalúa una competencia de acuerdo al estándar en un determinado nivel (vease Tabla 1), y en un contexto dado por los contenidos de ciencias y matemáticas del currículo oficial para la educación secundaria.

En la elaboración de un instrumento de evaluación es fundamental considerar las preguntas referidas a qué se evalúa; cuáles son las partes legítimamente implicadas en la evaluación; quién evalúa; cómo se evalúa; qué metodología utilizar, qué feedback se ofrece a las partes implicadas, planes de mejora generados por la evaluación, opinión de las partes implicadas sobre la evaluación (Muñiz & Fonseca-Pedrero, 2008).

El test para el diagnóstico de competencias científico-matemáticas está dirigido a matriculados en las carreras de Pedagogía en Biología, Pedagogía en Física, Pedagogía en Química y Pedagogía en Matemática, impartidas por la Universidad Metropolitana de Ciencias de la Educación (UMCE) a partir del año académico 2015, y pretende medir tres competencias y los niveles de logro que muestran tales estudiantes. Esto se hace con el fin de indagar en las potenciales "brechas" existentes, entendidas como diferencias en los niveles de logro de

la competencia evaluada entre cada estudiante, y así implementar un plan de nivelación que contribuya a cerrar esas brechas.

Este test solo mide aquellos indicadores factibles de ser evaluados en una prueba de lápiz y papel. Los ítems que se incluyen en el examen, son de selección múltiple.

Estructura del instrumento

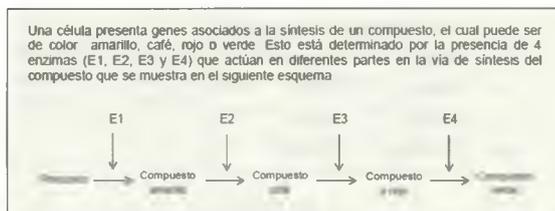
Se elaboró un instrumento en dos formas (F1 y F2) que evalúa el nivel de logro de cada competencia en cada uno de sus tres niveles de desempeño. Cada forma aplicada consideraba 60 ítems con estímulos contextualizados a los cuatro subsectores curriculares (matemática, física, química y biología) distribuidos de forma equitativa (15 ítems por subsector). Desde la perspectiva del rasgo a evaluar, de los 60 ítems, 20 evaluaban la competencia 1: "Razonamiento Lógico"; 20 evaluaban la competencia 2: "Modelamiento"; y otros 20 evaluaban la competencia 3: "Argumentación"

Cada ítem que conforma el test diagnóstico está constituido por un cuerpo de ítem y sus respectivas opciones de respuesta. El cuerpo del ítem considera un estímulo que debe ser presentado y una pregunta o afirmación referida al estímulo (imágenes 1, 2 y 3). El cuerpo del ítem se elabora siguiendo las siguientes recomendaciones generales: i) presenta un problema o situación elaborada, de manera tal que tenga sentido completo, esto es, que no sea necesaria más información que la que se está presentando para responder; ii) considera toda la información necesaria para comprender la tarea requerida y seleccionar la respuesta, sin que sea necesario recurrir a información adicional. Finalmente, iii) el enunciado debe ser gramaticalmente consistente con las opciones de respuesta y, en lo posible, debe evitar proposiciones negativas.

En cuanto a las opciones de respuesta, se consideran cuatro alternativas, tres de las cuales deben ser completamente incorrectas, claras y reflejar posibles errores de los postulantes (plausibles). La cuarta respuesta, correcta, debe ser precisa, completa e irrefutable. Los distractores en cada pregunta consideraron los siguientes criterios: i) se evitó incluir distractores que sean parcialmente correctos, salvo que la estructura y forma del ítem así lo requiera; ii) mantienen equilibrio y homogeneidad (incluyendo la respuesta correcta) en cuanto a: longitud, complejidad, estilo de redacción, sintaxis, morfología (género, número, persona, tiempo, modo); iii) de ser necesario incluir una proposición negativa, esta deberá aparecer al menos en dos distractores; iv) se debe evitar la repetición y/o sinonimia de términos o vocablos entre el cuerpo del ítem y la respuesta correcta. En caso de ser necesaria la repetición, deberá incluirse en más de

un distractor; y finalmente, los distractores iv) deben reflejar errores reales, posibles de cometer por los sujetos.

Figura 1: Ejemplo de ítem en el contexto de la biología. Este ítem evalúa la competencia 1 en el nivel 3, y muestra un índice de discriminación alto ($rb = 0,41$), y un grado de dificultad medio ($p = 48,5\%$) en el análisis posterior al pilotaje. A continuación, las Figuras 1, 2 y 3 muestran ejemplos del área de biología, química y de la física.

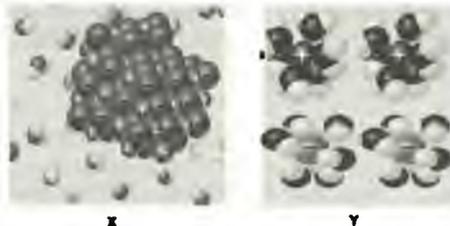


Si en un estudio in vitro se necesita obtener el compuesto rojo, ¿cuál de las opciones presenta los componentes necesarios para sintetizar este compuesto?

- A) Solo la enzima 1, la enzima 2 y el compuesto café.
- B) Solo la enzima 2, la enzima 3 y el precursor.
- C) Solo la enzima 2, la enzima 3 y el compuesto amarillo.
- D) Solo la enzima 1, el compuesto amarillo y el compuesto café.

Figura 2. Ejemplo de ítem en el contexto de la química. Este ítem evalúa la competencia 2 en el nivel y muestra un índice de discriminación alto ($rb = 0,37$), y un grado de dificultad medio ($p = 68,9\%$)

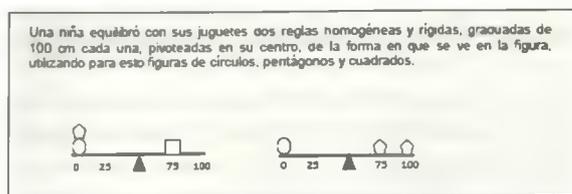
A continuación se presenta un modelo esquemático del estado de ionización del $\text{Al}(\text{OH})_3$, en donde las esferas (+) representan a cationes, las esferas (-) a aniones y las esferas blancas y grises a moléculas de agua



En relación al esquema y al gráfico presentado, es correcto afirmar que

- I) a pH 2 el estado de ionización de la sal estará representada esquemáticamente como en X
- II) a pH 12 el estado de ionización de la sal estará representada esquemáticamente como en Y.
- III) entre el rango de pH de 4 a 10 la sal estará representada esquemáticamente como en X.

- A) Solo I
- B) Solo I y III
- C) Solo II y III
- D) I, II y III



¿Cuál de las figuras tiene la menor y mayor masa, respectivamente?

	Menor masa	Mayor masa
A)	Círculo	Cuadrado
B)	Pentágono	Círculo
C)	Cuadrado	Círculo
D)	Pentágono	Cuadrado

= 80,58%)

Figura 3: Ejemplo de ítem en el contexto de la física. Este ítem evalúa la competencia 1 en el nivel 2, y muestra un índice de discriminación bajo ($rb = 0,27$), y un grado de dificultad bajo ($p = 80,58\%$).

Análisis de confiabilidad

Para el análisis de confiabilidad se utilizó el estadístico alpha de Cronbach. Este es un índice de consistencia interna que mide la homogeneidad de las preguntas promediando las correlaciones entre los ítems. Para este caso, el índice de confiabilidad alfa de Cronbach promedio supera el 0,9 ($\alpha > 0,9$), parámetro que indica alta confiabilidad. En lo que respecta a la confiabilidad por cada una de las competencias evaluadas, estas superaron el 0,7 ($\alpha > 0,7$).

Tabla 2: Confiabilidad y error de medida de las formas de prueba

Forma	Alfa de Cronbach	Error de medida
F1	0,921	1,67
F2	0,931	1,51
Promedio	0,926	1,59

Tabla 3: Confiabilidad y error de medida por competencia en F1

Forma 1	Alfa de Cronbach	Error de medida
Competencia 1	0,743	1,17
Competencia 2	0,777	1,28
Competencia 3	0,800	1,11

Tabla 4: Confiabilidad y error de medida por competencia en F2

Forma 2	Alfa de Cronbach	Error de medida
Competencia 1	0,736	1,17
Competencia 2	0,800	1,12
Competencia 3	0,856	0,94

Análisis de ítems

Para el análisis, desde la perspectiva de la Teoría Clásica de los Tests, se consideró:

Índice de dificultad (proporción de aciertos): el índice de dificultad para cada ítem (p) se define como la proporción de sujetos que responde correctamente al mismo del total de personas que abordó la pregunta (Muñiz, Fidalgo, Cueto, Martínez & Moreno, 2005). En la Tabla 5 se muestran los criterios para el análisis de la dificultad de los ítems.

Tabla 5: criterios para establecer puntos de corte para índice de dificultad.

p entre 0.76 y 0.90	Fácil
p entre 0.41 y 0.75	Media
p entre 0.10 y 0.40	Difícil

Índice de discriminación: se define como la correlación que se establece entre las puntuaciones que obtienen los sujetos en un ítem particular y las puntuaciones totales en el test (Muñiz et al., 2005). Se estimaron los índices de discriminación de los ítems por medio del *índice de correlación biserial* (rb). Este permite relacionar respuestas de tipo dicotómicas y discretas (acierto versus no acierto) con una escala de tipo continua (puntuación total sobre la escala o prueba), evaluando así el grado de asociación y, por extensión, de discriminación de los ítems. Los puntos de corte se establecieron de acuerdo a los criterios sugeridos por Muñiz et al (2005). En la tabla 6 se resumen los puntos de corte que se utilizaron para clasificar el índice de correlación biserial (rb).

Tabla 6: criterios para establecer puntos de corte para índice de discriminación.

rb	Clasificación del ítem
Igual o mayor que 0,40	Discrimina muy bien
Entre 0,30 y 0,39	Discrimina bien
Entre 0,20 y 0,29	Discrimina poco
Entre 0,10 y 0,19	Límite. Se debe mejorar
Menor de 0,10	Carece de utilidad para discriminar

Análisis de las opciones incorrectas o distractores: en un ítem se denomina *opción incorrecta* o *distractor* a las alternativas de respuesta falsas o posibilidades de respuesta incorrectas que este tiene. Como señalan Muñiz et al. (2005), es fundamental que todas las alternativas incorrectas incluidas, en tanto opciones de respuesta al ítem, resulten igualmente atractivas para las personas evaluadas que desconocieran la respuesta correcta. En la Tabla 7 se muestran los criterios usados para evaluar el comportamiento de los distractores.

Tabla 7: criterios para usados para valorar el comportamiento de los distractores.

Criterio 1	Deben ser elegidos por el 2% o más de los que abordaron la pregunta.
Criterio 2	Deben presentar un coeficiente de correlación biserial (rb) con el puntaje total negativo, o al menos inferior a 0,2 en la alternativa correcta.

En la Tabla 8 se muestran los resultados de los ítems que aprobaron el pilotaje. Los ítems descartados fueron 40 en total. De estos, la mayoría fueron descartados porque no discriminaban adecuadamente ($rb < 0,1$), o bien porque los distractores no cumplían con los criterios establecidos.

Tabla 8: Resumen de ítems que pasan la fase de validación del test en sus formas 1 y 2.

FORMA	PREG	N	p (%)	rb	FORMA	PREG	N	p (%)	rb
1	1	50	48,54	0,41	2	1	71	67,62	0,411
1	2	19	18,45	0,414	2	3	39	37,14	0,243
1	3	35	33,98	0,206	2	5	31	29,52	0,353
1	4	56	54,37	0,197	2	6	39	37,14	0,33
1	6	31	30,1	0,225	2	7	73	69,52	0,343
1	7	75	72,82	0,342	2	8	70	66,67	0,326
1	10	78	75,73	0,345	2	10	86	81,9	0,269
1	11	72	69,9	0,246	2	11	81	77,14	0,283
1	12	90	87,38	0,474	2	12	96	91,43	0,2
1	16	67	65,05	0,293	2	15	55	52,38	0,22
1	20	51	49,51	0,485	2	18	54	51,43	0,219
1	22	83	80,58	0,268	2	22	35	33,33	0,212
1	23	27	26,21	0,415	2	26	18	17,14	0,32
1	24	91	88,35	0,327	2	27	45	42,86	0,22
1	25	22	21,36	0,221	2	28	65	61,9	0,257
1	27	29	28,16	0,328	2	29	43	40,95	0,507
1	29	34	33,01	0,201	2	30	20	19,05	0,452
1	32	71	68,93	0,376	2	31	39	37,14	0,321
1	33	66	64,08	0,471	2	32	43	40,95	0,214
1	35	63	61,17	0,414	2	34	25	23,81	0,32
1	36	57	55,34	0,466	2	35	64	60,95	0,415
1	38	36	34,95	0,288	2	36	60	57,14	0,512
1	40	67	65,05	0,385	2	37	52	49,52	0,285
1	41	53	51,46	0,234	2	38	49	46,67	0,447
1	42	25	24,27	0,286	2	40	85	80,95	0,343
1	44	44	42,72	0,523	2	41	49	46,67	0,426
1	45	49	47,57	0,457	2	42	75	71,43	0,67
1	47	51	49,51	0,319	2	43	47	44,76	0,214
1	49	79	76,7	0,335	2	44	52	49,52	0,261
1	50	47	45,63	0,439	2	47	74	70,48	0,452

FORMA	PREG	N	p (%)	rb	FORMA	PREG	N	p (%)	rb
1	51	59	57,28	0,54	2	49	86	81,9	0,395
1	52	70	67,96	0,324	2	50	59	56,19	0,284
1	53	64	62,14	0,214	2	51	30	28,57	0,517
1	56	78	75,73	0,33	2	52	68	64,76	0,39
1	57	67	65,05	0,315	2	53	17	16,19	0,36
1	58	63	61,17	0,389	2	54	25	23,81	0,467
1	59	75	72,82	0,458	2	55	70	66,67	0,348
1	60	73	70,87	0,48	2	56	72	68,57	0,297
					2	57	51	48,57	0,569
					2	58	78	74,29	0,448
					2	59	79	75,24	0,312
					2	60	69	65,71	0,477

Conclusiones

Al considerar los criterios de análisis de los ítems de ambas formas del test piloto, se logró la aprobación de 38 de los 60 ítems en la forma 1 y de 42 de los 60 ítems en la forma 2 (véase Tabla 7). De aquí que el test final para el diagnóstico de competencias científico-matemáticas quedó conformado por 80 ítems validados.

El test construido a partir de las preguntas que superaron la revisión y validación de los ítems es considerado de una alta confiabilidad ($\alpha > 0,9$). La confiabilidad asociada a la medición de las competencias científico-matemáticas evaluadas es también alta ($\alpha > 0,7$).

El índice de dificultad p del test indica un promedio próximo al 54%, con una desviación estándar del 19%. Esta estimación se obtuvo a partir del índice de dificultad de cada ítem aprobado, y permite afirmar que el test posee un grado de dificultad medio en la población testeada. Debe destacarse, no obstante, que el índice de dificultad no es una constante entre grupos.

El índice de discriminación rb promedio del test se encuentra en torno al 0,35. Este parámetro, asociado a los criterios previamente establecidos, permite sostener que el test "discrimina bien".

Por razones de espacio no se incluye el test final validado.

Por último, la principal ventaja del uso de este tipo de instrumento es que constituye un modelo de evaluación estandarizado y ampliamente utilizado, de fácil aplicación y resultados rápidos, y dirigible a la toma de decisiones a la hora de aplicar un plan de nivelación de competencias científico-matemáticas.

Proyecciones

Los resultados obtenidos otorgan información que permite determinar los "puntajes de corte" en los niveles

de logro de las competencias de los estudiantes. En esta misma línea, actualmente se trabaja con un método centrado en el test, específicamente el método de Angoff (1984), para intentar dar cuenta de este desafío, y con ello mejorar la interpretación de los resultados del test.

Agradecimientos

El equipo agradece a la Universidad Metropolitana de Ciencias de la Educación (UMCE) y a MECESUP el posibilitar la ejecución del proyecto UMC-1302.

Referencias

Adúriz-Bravo, A. (2012). Competencias metacientíficas escolares dentro de la formación del profesorado de ciencias, en Badillo, E., García, L., Marbà, A. y Briceño, M. (coords.). *El desarrollo de competencias en la clase de ciencias y matemáticas*, 43-67. Mérida: Universidad de Los Andes.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational testing service.

Beneitone, P., César, E., González, J., Maleta, M., Siufi, R., & Wagenaar, G. (2007). *Reflexiones y perspectivas de la educación superior en América Latina. Informe Final Proyecto Tuning América Latina*. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Reflexiones+y+perspectiva+s+de+la+Educaci+n+Superior+en+Am+rica+Latina+Informe#>

Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5(3), 521-551.

Downing, S. M., & Haladyna, T. M. (2006). *Handbook of Test Development*. Lawrence Erlbaum Associates Publishers. <http://doi.org/10.1080/15305050701813433>

Eurydice. (2002). *Competencias clave*. Retrieved from [http://www.gobiernodecanarias.org/educacion/Portal/WEBicec/docs/pcb/competencias_clave\(EURIDYCE\).pdf](http://www.gobiernodecanarias.org/educacion/Portal/WEBicec/docs/pcb/competencias_clave(EURIDYCE).pdf)

García-Cartagena, Y., & Reyes, D. (2012). Robótica educativa y su potencial mediador en el desarrollo de las competencias asociadas a la alfabetización científica. *Educación y Tecnología*, 2, 42-55.

MINEDUC. (2009). *Objetivos fundamentales y contenidos mínimos obligatorios de la Educación Básica y Media*. Retrieved from http://www.curriculumlineamineduc.cl/605/articles-34641_bases.pdf

MINEDUC. (2012). *Estándares orientadores para carreras de pedagogía en educación media*. Retrieved from <http://www.mineduc.cl/usuarios/cpeip/doc/201206011651510.LibroEstAndaresEducacionMedia.pdf>

Monereo, C., & Pozo, J. I. (2007). Competencias para convivir con el siglo XXI. *Cuadernos de Pedagogía*, 1(370), 12-18.

Muñiz, J., Fidalgo, A. M., Cueto, G., Martínez, R., & Moreno, R. (2005). *Análisis de los ítems*. Madrid: La Muralla.

Muñiz, J., & Fonseca-Pedrero, E. (2008). Construcción de Instrumentos de Medida para la Evaluación Universitaria. *Revista de Investigación En Educación*, 5, 13-25. Retrieved from <http://webs.uvigo.es/reined/>

OCDE. (2002). Definition and Selection of Competences (DeSeCo) Theoretical and Conceptual Foundations, 27. <http://doi.org/10.1007/s13398-014-0173-7.2>

OCDE. (2006). *PISA 2006. Marco de la evaluación*. Retrieved from <http://www.oecd.org/pisa/39732471.pdf>

Reyes, D., & García-Cartagena, Y. (2014). Desarrollo de habilidades científicas en la formación inicial de profesores de ciencias y matemáticas. *Educación Y Educadores*, 17(2), 271-285. <http://doi.org/10.5294/edu.2014.17.2.4>

UMCE. (2013). Plan de Mejoramiento para la nivelación de competencias científicas aplicado a estudiantes de primer año de Formación Inicial Docente de la Facultad de Ciencias Básicas basado en el modelo de Indagación Científica. Retrieved from http://www.mecesup.cl/index2.php?id_seccion=4957&id_portal=59&id_contenido=28900