



Universidad
Inca Garcilaso de la Vega
Nuevos Tiempos. Nuevas Ideas

FACULTAD DE PSICOLOGÍA Y TRABAJO SOCIAL

ELABORACIÓN DE INSTRUMENTOS
ESTANDARIZADOS PARA LA MEDICIÓN DE LA
COMPRENSIÓN LECTORA EN ESTUDIANTES DE
EDUCACIÓN BÁSICA

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el título profesional de LICENCIADO EN PSICOLOGÍA

AUTOR

Bachiller: ALTAMIRANO IZQUIERDO, ALFREDO JESUS

ASESOR

Mg. Liñán Colchado, Gloria María

Lima - Perú

2022

DEDICATORIA

A mi abuela, Juanita Cruz Dioses, ejemplo de aprendizaje, persistencia y amor.



AGRADECIMIENTOS

Agradezco a la Sra. Liliana Miranda Molina, ex Jefa de la Unidad de Medición de la Calidad Educativa, por esta muy enriquecedora experiencia profesional. Asimismo, agradezco a la Dra. Gloria Liñán Colchado por sus acuciosos aportes en la elaboración del presente documento.



RESUMEN

La experiencia laboral presentada y descrita en este documento se realizó en la Oficina de Medición de la Calidad de los Aprendizajes (UMC) del Ministerio de Educación. En dicha institución pertencí al Equipo de Evaluación y me encargué del diseño y elaboración de pruebas estandarizadas de comprensión lectora dirigidas a estudiantes de distintos grados de primaria y secundaria a nivel nacional, como la Evaluación Censal de Estudiantes (ECE). Asimismo, esta experiencia implicó aprender a analizar reportes psicométricos elaborados por los estadísticos de la Oficina a fin de hacer los ajustes necesarios en las pruebas elaboradas. En algunos excepcionales, implicó también familiarizarse y aplicar procedimientos para una codificación confiable de las respuestas a ítems de respuesta abierta (un tipo particular de ítem algunas veces incluido en estas pruebas). Esta experiencia profesional permitió conocer a fondo el proceso real de realización de procesos de evaluación estandarizada a gran escala, aplicar y profundizar los conocimientos psicológicos que ya se tenían sobre el tema, y adquirir experiencia inicial en el campo de la psicometría y la construcción de tests específicos para medir la comprensión lectora.

Palabras clave: **prueba estandarizada, ítem, comprensión lectora, psicometría, constructo.**

ABSTRACT

The professional experience described in this document was held at the Ministry of Education's Office of Measurement of Learning Quality (UMC). At this institution, I worked as a member of the Assessment Team, responsible of tasks such as the design and elaboration of reading comprehension standardized tests aimed at students from different grades in primary and secondary nationwide (e.g., the National Student Assessment – ECE). In addition, I had to learn how to analyze psychometric reports prepared by the Office's statisticians, in order to make adjustments to trialed items. In exceptional cases, this work also implied getting to know and applied procedures to ensure a reliable coding of student responses to open-constructed items (a particular type of item sometimes included in these tests). This professional experience allowed me to deeply understand the process of large-scale standardized-test construction, to apply and expand the psychologic knowledge I already had, and to acquire initial experience in the field of psychometric and item construction to measure reading comprehension.

Keywords: standardized test, item, reading comprehension, psychometrics, construct.

ÍNDICE

DEDICATORIA	2
AGRADECIMIENTOS	3
RESUMEN	4
ABSTRACT	5
ÍNDICE	6
ÍNDICE DE FIGURAS	8
INTRODUCCIÓN	9
CAPÍTULO I: ASPECTOS GENERALES	11
1.1. DESCRIPCIÓN DE LA EMPRESA O INSTITUCIÓN	11
1.2. DESCRIPCIÓN DEL PRODUCTO O SERVICIO	12
1.3. UBICACIÓN GEOGRÁFICA Y CONTEXTO SOCIOECONÓMICO	13
1.4. ACTIVIDAD GENERAL O ÁREA DE DESEMPEÑO	14
1.5. MISIÓN Y VISIÓN	15
CAPITULO II: DESCRIPCIÓN GENERAL DE LA EXPERIENCIA	17
2.1. ACTIVIDAD PROFESIONAL DESARROLLADA	17
2.2. PROPÓSITO DEL PUESTO Y FUNCIONES ASIGNADAS	18
CAPITULO III: FUNDAMENTACIÓN DEL TEMA ELEGIDO	21
3.1. TEORÍA Y LA PRÁCTICA EN EL DESEMPEÑO PROFESIONAL	21
3.1.1. Medición en Psicología	21

3.1.2. Evaluación por criterios	22
3.1.3. Modelo psicométrico para el diseño de las pruebas y análisis de resultados	22
3.1.4. Validez y confiabilidad de las pruebas	23
3.1.5. Constructos evaluados	24
3.2. TEORÍA Y LA PRÁCTICA EN EL DESEMPEÑO PROFESIONAL	25
3.2.1. Revisión y alineación de las pruebas con el documento curricular vigente	25
3.2.2. Ítems de la prueba	26
3.2.3. Codificación de respuestas	29
3.2.4. Puntos de corte	30
3.2.5. Reporte de resultados	32
CAPITULO IV: PRINCIPALES CONTRIBUCIONES	33
RECOMENDACIONES	33
REFERENCIAS BIBLIOGRÁFICAS	35
ANEXO	36

ÍNDICE DE FIGURAS

FIGURA 1. Organigrama de la Secretaría de Planificación Estratégica del MINEDU	11
FIGURA 2. Edificio de la Oficina de Medición de la Calidad de los Aprendizajes	14
FIGURA 3. Currículo Nacional de Educación Básica (2016)	24
FIGURA 4. Ejemplo de ítem de respuesta abierta	27
FIGURA 5. Ejemplo de ítem de opción múltiple	28
FIGURA 6. Ejemplo de códigos de un ítem de respuesta abierta	30
FIGURA 7. Niveles de desempeño de la prueba ECE (Evaluación Censal de Estudiantes) de segundo grado de primaria en Comprensión lectora	31
FIGURA 8. Portada de reporte de resultados de la prueba ECE (Evaluación Censal de Estudiantes) de segundo grado de primaria dirigido a docentes	32

INTRODUCCIÓN

La Oficina de Medición de la Calidad de los Aprendizajes (UMC) es el órgano del Ministerio de Educación encargado de realizar evaluaciones periódicas del rendimiento estudiantil a fin de proveer de información a los tomadores de decisiones a todo nivel (nacional, regional, local y las propias escuelas) para el diseño e implementación de políticas educativas adecuadas. El presente documento describe la experiencia profesional ejercida en esta institución, muy vinculada al terreno psicológico, especialmente al de la psicometría, la cual permite enriquecer conocimientos y vincularlos con la práctica. Respecto de esto último, podría considerarse como una experiencia privilegiada en tanto que muy pocas instituciones en el Perú elaboran pruebas estandarizadas de rendimiento estudiantil con grandes muestras (pruebas a gran escala) y solo la UMC, además, tiene contacto directo con algunas etapas en los procesos de desarrollo y aplicación de pruebas internacionales que se aplican en nuestro país (PISA, ICCS, UNESCO-Llece). En las páginas a continuación se presentarán todos los pormenores de esta experiencia, específicamente en el terreno de la construcción de pruebas de comprensión lectora.

En el primer capítulo se presentará la institución en la que esta experiencia profesional fue realizada. Esto incluye no solo su ubicación geográfica, sino también su posición institucional dentro del organigrama del Ministerio de Educación. Acto seguido, en el segundo capítulo, se describirá el trabajo realizado: funciones encomendadas, descripción del puesto asignado y procesos seguidos. En el tercer capítulo, se desarrollará los principales conceptos teóricos que guiaron la práctica profesional realizada, tales como los modelos psicométricos que usa la institución para analizar el funcionamiento de los ítems y para reportar los resultados obtenidos por los estudiantes, además del referente tomado para la definición de los constructos a evaluar, considerando que estos procesos de medición están enmarcados dentro de las políticas educativas. Asimismo, en este capítulo se explicarán los procesos realizados en cada una de las varias etapas de este proceso, desde la concepción y diseño de las pruebas hasta la publicación de los resultados. Finalmente, el documento incluye una lista de conclusiones respecto de la experiencia llevada a cabo, así como

recomendaciones que ayudarían a potenciar el trabajo y el impacto de esta institución tanto en el desarrollo de la psicometría en el Perú como en la toma de decisiones para la mejora de los aprendizajes de los estudiantes peruanos.

CAPÍTULO 1: ASPECTOS GENERALES

1.1. DESCRIPCIÓN DE LA EMPRESA O INSTITUCIÓN

La actividad profesional se realizó en el Ministerio de Educación, en la Oficina de Medición de la Calidad de los Aprendizajes (UMC), antes denominada Unidad de Medición de la Calidad Educativa, que forma parte de la Secretaría de Planificación Estratégica de dicho Ministerio, ente de especial importancia en este Ministerio, casi al nivel de los dos Viceministerios que lo conforman. Es una de las cuatro oficinas que integran dicha Secretaría, junto con la Oficina de Planificación Estratégica y Presupuesto, la Oficina de Seguimiento y Evaluación Estratégica, y la Oficina de Tecnologías de la Información y Comunicación.

Figura 1. Organigrama de la Secretaría de Planificación Estratégica del MINEDU



Fuente: http://www.minedu.gob.pe/p/xtras/organigrama_minedu.pdf

Según su página web, la Oficina de Medición de la Calidad de los Aprendizajes, tiene como objetivo institucional “producir y difundir información relevante y confiable sobre los resultados de las evaluaciones y sus factores asociados, para contribuir a la toma de

decisiones en las instancias de decisión política educativa, la comunidad educativa y la sociedad en general, con el fin de mejorar la calidad del sistema educativo de nuestro país” (<http://umc.minedu.gob.pe/la-umc/>).

La Oficina de Medición de la Calidad de los Aprendizajes fue creada a fines de la década del 90 y cuenta en la actualidad con un total aproximado de 60 trabajadores. Se conforma en su interior de cuatro grandes equipos o unidades:

- Equipo de Evaluación: Encargado de diseñar las pruebas y los reportes de resultados a partir de los resultados obtenidos por los estudiantes en ellas.
- Equipo de Factores Asociados: Encargado de los procedimientos estadísticas para la definición de la muestra, el análisis de la idoneidad de los ítems y el procesamiento de los resultados obtenidos por los estudiantes.
- Equipo de Planificación: Encargado de diseñar el operativo de aplicación de las pruebas a nivel nacional y de realizar las coordinaciones respectivas para ello.
- Equipo de Administración: Encargado de llevar a cabo los procesos administrativos necesarios para el funcionamiento interno de la Oficina (contratos de trabajadores, compras de materiales, etc.).

Cada uno de estos equipos cuenta con un Coordinador, los cuales son dirigidos por el Jefe de la Oficina. En el periodo en el que laboré en esta institución (2007-2011), la Jefa a cargo era la Sra. Liliana Miranda Molina.

1.2. DESCRIPCIÓN DEL PRODUCTO O SERVICIO

La Oficina de Medición de la Calidad de los Aprendizajes tiene a su cargo el diseño, aplicación, supervisión y reporte de resultados de evaluaciones estandarizadas dirigidas a estudiantes de educación básica (Inicial, Primaria y Secundaria) a nivel nacional (tanto colegios públicos como privados).

Asimismo, la Oficina de Medición de la Calidad de los Aprendizajes se encarga de aplicar - no de diseñar- algunas pruebas de carácter internacional, como las pruebas PISA¹, ICCS o

¹ La Prueba PISA es elaborada por la OCDE (Organización para la Cooperación y Desarrollo Económico), se aplica cada tres años en decenas de países y evalúa tres dominios en estudiantes de quince años: Comprensión lectora, Matemáticas y Ciencias. Por su parte, la Prueba ICCS es elaborada por la IEA (Asociación Internacional para la Evaluación del Logro Educativo), se aplica en decenas de países y evalúa competencias ciudadanas en estudiantes de segundo grado de secundaria. Finalmente, el Laboratorio de la UNESCO evalúa

las de Unesco, las cuales sirven para comparar los resultados de nuestro país con los de otros países participantes en dichas pruebas. Dichas pruebas son elaboradas por agencias internacionales especializadas, por lo que los países participantes reciben las pruebas listas para su aplicación en la muestra de estudiantes. Cuando estas pruebas vienen elaboradas en un idioma diferente al castellano (por ejemplo, inglés en el caso de PISA e ICCS), la Oficina de Medición de la Calidad de los Aprendizajes se encarga, además, de traducir dichas pruebas al castellano peruano siguiendo procedimientos estandarizados.

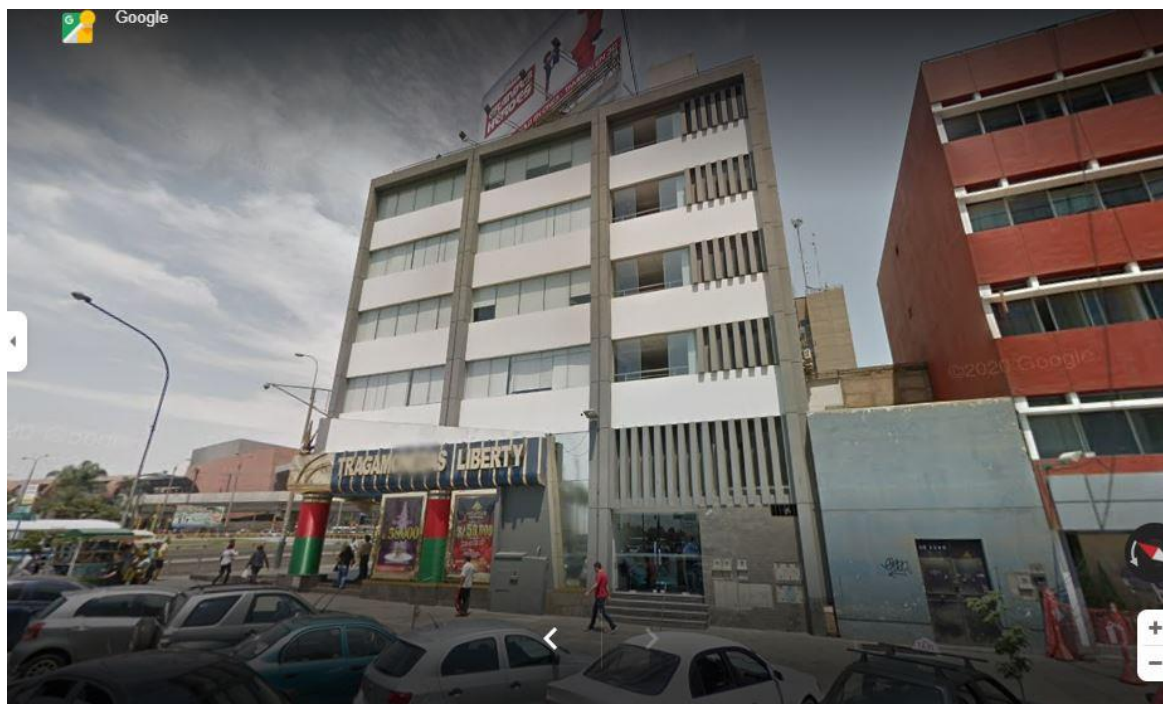
Por último, la Oficina de Medición de la Calidad de los Aprendizajes tiene también a su cargo la elaboración de los reportes de resultados obtenidos en la prueba. Dichos reportes suelen estar dirigidos a diverso público; se elaboran reportes diferenciados, en información y lenguaje, según cada destinatario: padres de familia, docentes, directores, autoridades educativas (en el caso de los reportes dirigidos a los docentes y padres de familia, estos son enviados en físico a los directores de las escuelas quienes tienen, a su vez, la obligación de reenviarlos a sus reales destinatarios). Estos reportes suelen contener, además de los resultados de los estudiantes, recomendaciones de práctica y política educativa a partir precisamente de los resultados obtenidos.

1.3. UBICACIÓN GEOGRÁFICA Y CONTEXTO SOCIOECONÓMICO

La Oficina de Medición de la Calidad de los Aprendizajes se ubica en la ciudad de Lima, en un edificio de tres pisos (de uso exclusivo) en la calle Morelli 109, San Borja, a pocas cuadras de la sede central del Ministerio de Educación. Está ubicado en un vecindario de clase media, rodeados de casas residenciales, edificios multifamiliares, oficinas y negocios varios. También podemos encontrar en las inmediaciones dos colegios públicos, en los que a veces se implementan algunos operativos piloto dada su cercanía. Por otro lado, es una zona con múltiples vías de acceso y transitando por diversas líneas de transporte público, incluyendo la Estación La Cultura del Metro de Lima, ya que está ubicada entre las avenidas Canadá, Javier Prado, Aviación y San Luis.

a estudiantes de tercer y sexto grado de primaria en casi todos los países de América Latina y el Caribe en Comprensión lectora, Producción escrita, Matemáticas y Ciencia.

Figura 2. Frontis del edificio de la Oficina de Medición de la Calidad de los Aprendizajes



1.4. ACTIVIDAD GENERAL O ÁREA DE DESEMPEÑO

La Oficina de Medición de la Calidad de los Aprendizajes tiene a su cargo llevar a cabo procesos de evaluación estandarizada de los aprendizajes de estudiantes de educación básica a nivel nacional (incluyendo a estudiantes que tienen una lengua indígena como lengua materna). Ello implica las siguientes etapas:

- Diseño del marco teórico de la prueba: Se definen las variables a medir (por ejemplo, comprensión lectora, alfabetización científica, etc.) y el marco teórico a partir del cual se las está conceptualizando (por ejemplo, el enfoque comunicativo, en el caso de Comprensión lectora o Producción escrita).
- Diseño del marco muestral: Se define la población a evaluar, así como la muestra, estimando el alcance descriptivo que tendrá la prueba. Por lo general, las mediciones de la Oficina de Medición de la Calidad de los Aprendizajes contemplan muestras nacionales con estratos público-privado y urbano-rural.
- Elaboración de la prueba: Trabajo “de escritorio” en el que el equipo de Evaluación se dedica a la elaboración de los reactivos e ítems que contemplará la prueba. Por lo general, se elaboran muchos más ítems de los requeridos, a fin de que pasen por el filtro de la

siguiente etapa.

- Aplicación piloto: Se aplican las pruebas en una muestra de menor tamaño que la muestra definitiva/final, pero manteniendo cierta globalidad nacional. El objetivo es identificar (y descartar) ítems que resultan confusos para los estudiantes debido a su formulación o que no corresponden al nivel de dificultad esperado para el grado a evaluar. Asimismo, ítems que no llegan a ser descartados pueden de todos modos ser mejorados a partir de los *insights* que brinda este proceso.

- Elaboración de prueba definitiva: Se retiran los ítems problemáticos de la prueba, se ajustan otros que necesitan o mejorar, o incluso, aunque no es lo deseable, se elaboran nuevos ítems en caso se necesite evaluar algún aspecto de la variable para la cual no sobrevivió ningún ítem.

- Aplicación definitiva: Implica el diseño y la implementación de una serie de procesos rigurosos que apuntan a garantizar la confiabilidad de la información recogida (velar por la confidencialidad de las pruebas en todo su traslado hasta las escuelas, evitar las copias, asegurar que el tiempo no sea un obstáculo para que los estudiantes respondan la prueba, etc.).

- Análisis psicométrico: A partir de una serie de procedimientos estadísticos, se establecen los niveles de dificultad de la prueba y se producen los resultados a partir de la información consignada por los estudiantes. Previamente, implica también un proceso de depuración de la información, en el que se detectan y descartan respuestas en las que se marcó más de una alternativa de respuesta (si se trata de una pregunta de opción múltiple), preguntas sin responder, etc.

- Reporte de resultados: Con la información obtenida y procesada se pasa a elaborar reportes dirigido a distinto público en los que se muestran los resultados y se ensaya una descripción de lo que dichos resultados significan en términos de logro de aprendizaje por parte de los estudiantes. Adicionalmente, estos reportes suelen tener recomendaciones para elevar dichos niveles de aprendizaje.

1.5. MISIÓN Y VISIÓN

Formalmente, la Oficina de Medición de la Calidad de los Aprendizajes no cuenta con una

misión y una visión oficialmente establecidas; sin embargo, a partir de la labor que realiza, podríamos decir lo siguiente:

Misión: Retroalimentar al sistema educativo peruano con información de calidad respecto al estado actual y el progreso de los aprendizajes de los estudiantes peruanos a fin de orientar la mejor toma de decisiones en materia de política educativa

Visión: Convertirse en un referente nacional e internacional de medición de aprendizajes a partir de pruebas estandarizadas dirigidas a todas las áreas curriculares y grados de la educación básica en el Perú.

CAPITULO II: DESCRIPCIÓN GENERAL DE LA EXPERIENCIA

2.1. ACTIVIDAD PROFESIONAL DESARROLLADA

La actividad profesional se desarrolló en el Equipo de Evaluación de la Oficina de Medición de la Calidad de los Aprendizajes entre los años 2007 y 2011, en el cual se pudo aplicar y expandir conocimientos sobre medición psicológica y construcción de pruebas. Este equipo es el responsable del diseño del marco teórico de las evaluaciones y los instrumentos de evaluación en las distintas áreas curriculares para las que contaba con especialistas (hasta el año 2011, la Oficina de Medición de la Calidad de los Aprendizajes elaboraba pruebas fundamentalmente en las áreas de Comunicación -tanto para estudiantes hispanohablantes como para estudiantes de cuatro lenguas indígenas- y Matemáticas; en los años posteriores ampliaron su cobertura a las áreas de Ciencia, Tecnología y Ambiente, y de Ciencias Sociales y Ciudadanía). Asimismo, en un primer momento este Equipo también participaba activamente en la organización e implementación del proceso de aplicación de los instrumentos de evaluación a nivel nacional, considerando que casi todas las pruebas se aplicaban a nivel nacional (y siendo una de ellas una prueba censal, es decir, dirigida a todos los estudiantes del país en determinados grados), y de la elaboración de los manuales dirigidos a los involucrados en llevar a cabo dicha aplicación de la manera adecuada (desde los aplicadores contratados para este fin hasta los docentes y directores, para que tengan claro cuál era su rol durante la aplicación, pasando por los distintos funcionarios del operador logístico asignado a este proceso).

Respecto del procesamiento estadístico de la información y la validación y estandarización de los instrumentos a ser aplicados, esta era una labor cuya dirección recaía en el Equipo de Análisis y Factores Asociados; sin embargo, el Equipo de Evaluación realizaba coordinaciones constantes con ellos e incorporaba los ajustes necesarios en los instrumentos a partir de los reportes psicométricos que el Equipo de Análisis nos entregaba. Es decir, el proceso de elaboración de los instrumentos y de reporte de resultados finales seguía la siguiente secuencia:

- Elaboración del marco de la evaluación (Equipo de Evaluación)
- Elaboración de los instrumentos de evaluación (pruebas) a ser piloteadas (Equipo de Evaluación)

- Análisis psicométrico del funcionamiento de los ítems de las pruebas post aplicación piloto (Equipo de Análisis y Factores Asociados)
- Revisión del reporte psicométrico post aplicación piloto e incorporación de ajustes a los ítems identificados como mejorables (o, de ser necesario, elaboración de nuevos ítems, aunque como última opción, ya que no había posibilidades de realizar una nueva aplicación piloto) para obtener la versión final de las pruebas (Equipo de Evaluación)
- Análisis psicométrico de los ítems post aplicación definitiva a fin de ubicarlos en una escala creciente de dificultad (Equipo de Análisis y Factores Asociados)
- Establecimiento de puntos de corte en dicha escala de dificultad a fin de establecer los niveles de desempeño de las competencias evaluadas (Equipo de Evaluación)
- Procesamiento de los resultados obtenidos por los estudiantes en las pruebas (la Oficina de Medición de la Calidad de los Aprendizajes solo reporta hasta la fecha resultados a nivel de estudiante, nunca a nivel de institución educativa, como sí es el caso de algunas entidades evaluativas de otros países, como el SIMCE de Chile), ubicándolos en alguno de los niveles de desempeño previamente establecidos (Equipo de Análisis y Factores Asociados)
- Elaboración de los reportes finales de resultados dirigidos a docentes, directores, padres de familia y autoridades locales y regionales.

2.2. PROPÓSITO DEL PUESTO Y FUNCIONES ASIGNADAS

Como se señaló anteriormente, fui contratado como especialista del Equipo de Evaluación para el área curricular de Comunicación. El propósito de mi puesto era el de contribuir, en trabajo de equipo, a elaborar instrumentos de medición válidos y confiables en las competencias del área curricular de Comunicación, específicamente de Comprensión lectora y, en menor medida, de Producción de textos, dirigidos a estudiantes de primaria y secundaria.

En tal sentido, me fueron asignadas las siguientes funciones:

- Revisión de material bibliográfico y diversas experiencias de medición de la comprensión lectora y la producción escrita en estudiantes escolares.
- Elaboración de ítems de Comprensión lectora y Producción de textos para estudiantes

de primaria y secundaria.

- Elaboración de manuales dirigidos al personal a cargo de los procesos de aplicación de las pruebas.
- Coordinaciones logísticas y operativas con dicho personal.
- Coordinación de los procesos de codificación de las respuestas a las preguntas abiertas de las pruebas.
- Redacción de los reportes de resultados dirigidos a diversos públicos (docentes, directores).

Adicionalmente, me fue asignada la coordinación del proceso de implementación de la Prueba PISA 2009 en el Perú, dado el énfasis en Comprensión lectora que tenía aquella edición de esta prueba internacional y el uso del inglés como idioma de contacto. Cabe señalar que, en el caso de la prueba PISA, no cumplí funciones de elaboración de los ítems, ya que esta prueba es elaborada por los especialistas de la OECD (Organización para la Cooperación y el Desarrollo Económico, por sus siglas en inglés), de manera que mis funciones iniciaban más bien en la traducción de las pruebas hasta la aplicación de las mismas (tampoco participábamos en el análisis psicométrico de los ítems ni en el procesamiento de los resultados de los estudiantes). Específicamente realicé las siguientes actividades:

- Traducción del inglés al castellano de los ítems de Comprensión lectora de la prueba PISA 2009 y retraducción del castellano al inglés, para verificación del equipo técnico de la OECD.
- Supervisión de la traducción del inglés al castellano de los ítems de Matemáticas y Ciencias de la prueba PISA 2009 y retraducción del castellano al inglés, para verificación del equipo técnico de la OECD.
- Organización del proceso de aplicación de la prueba a nivel nacional en sus fases piloto y definitiva.
- Participación en el entrenamiento del equipo técnico de la OECD a los responsables de cada país de coordinar la codificación de las respuestas a las preguntas abiertas de las pruebas.

- Coordinación del proceso de codificación de las respuestas a las preguntas abiertas correspondiente a la sección de Comprensión lectora de la prueba.
- Realización de charlas y talleres informativos dirigidos a directores, docentes, autoridades locales y nacionales, y especialistas sobre las características de la prueba y sus implicancias en el trabajo de aula y las políticas educativas.

CAPITULO III: FUNDAMENTACIÓN DEL TEMA ELEGIDO

3.1. TEORÍA Y LA PRÁCTICA EN EL DESEMPEÑO PROFESIONAL

Una de las disciplinas con mayor desarrollo dentro de la evolución de la Psicología es el de la psicometría. En efecto, el interés por construir pruebas que permitan evaluar acertadamente diversos constructos de la mente y experiencia humanas han llevado a la construcción de una gran cantidad de tests, inventarios, escalas y demás instrumentos. A continuación, describiré algunos principios teóricos básicos que orientan la elaboración de tests, los cuales sirvieron de referencia para la experiencia laboral desempeñada en la Oficina de Medición de la Calidad de los Aprendizajes.

3.1.1. Medición en Psicología

Como señala Martínez Arias (2005), “los constructos psicológicos son abstracciones que únicamente pueden evaluarse indirectamente” (p.29), por lo que es importante llevar a cabo distintos procesos que garanticen la validez de los instrumentos a través de los cuales realizaremos su medición. Siguiendo a este mismo autor, los tests psicológicos se pueden clasificar en función de diversos criterios; sin embargo, para efectos de explicar el tipo de pruebas elaboradas por la Oficina de Medición de la Calidad de Aprendizaje, podemos emplear la siguiente clasificación:

- Rendimiento
- Inteligencia y aptitudes
- Personalidad, actitudes, intereses

Las pruebas diseñadas y aplicadas por la Oficina de Medición de la Calidad de los Aprendizajes son de rendimiento, ya que se refieren a “adquisiciones de aprendizaje o conocimientos específicos, con finalidades de diagnóstico y/o pronóstico concretas” (Martínez Arias, 2005, p.33). Cuando la noción de “rendimiento” es llevada al terreno educativo, se la puede entender como el resultado del proceso de enseñanza-aprendizaje (Silva Moreno, 2003). En el caso del sistema educativo peruano, dicho proceso de enseñanza-aprendizaje está orientado, en todas las escuelas, por lo planteado en el Diseño Curricular de Educación Básica – CNEB (2016).

3.1.2. Evaluación por criterios

Las pruebas diseñadas por la Oficina de Medición de la Calidad de los Aprendizajes se enmarcan dentro de lo que se denomina “evaluación por criterios”, cuyo objetivo es “determinar el estatus actual de rendimiento o conducta de un sujeto, o una clara descripción de sus características, normalmente con objeto de clasificarlo en una categoría determinada” (Martínez Arias, 2005). En ese sentido, una prueba de rendimiento referida a criterios sirve para “evaluar el estatus absoluto del sujeto con respecto a algún dominio de conductas bien definido” (Popham, 1978). Dichas conductas son establecidas en criterios objetivos formulados a partir de las definiciones operativas y ejes de los constructos a evaluar, los cuales se encuentran en el Currículo Nacional de Educación Básica del Ministerio de Educación (2016).

3.1.3. Modelo psicométrico para el diseño de las pruebas y análisis de resultados

La Oficina de Medición de la Calidad de los Aprendizajes, concretamente el Equipo de Análisis y Factores Asociados, basa sus análisis psicométricos y estadísticos en el modelo Rasch, uno de los varios modelos de la denominada Teoría de Respuesta al Ítem (TRI), la cual supera a la Teoría Clásica de los Test (TCT). Martínez Arias (2005), señala que la Teoría de Respuesta al Ítem “intenta dar una fundamentación probabilística al problema de la medición de rasgos y constructos no observables (...) [y] deriva de la consideración del ítem como unidad de análisis básica del test, en lugar de las puntuaciones totales, como era habitual en la Teoría Clásica de los Test” (p.238). Esta autora señala las siguientes limitaciones de la Teoría Clásica de los Test, superadas por la Teoría de Respuesta al Ítem (pp.238-239):

- 1) Los estadísticos del ítem de la Teoría Clásica de los Test dependen de la tendencia central y de la variabilidad del rasgo en la muestra y en la población.
- 2) Dependencia de los estadísticos del test de la población y/o muestra en que se calculen.
- 3) Dificultades prácticas con el concepto de paralelismo de las medidas.
- 4) La puntuación empírica total, base de la mayor parte de las inferencias de la Teoría Clásica de los Test, depende de los ítems concretos que componen el test y, normalmente, esta puntuación será más alta si los ítems son fáciles y más baja, si son difíciles.
- 5) El estimador usual de la puntuación verdadera, V' , depende del grupo de referencia.

Es posible rescatar como principal particularidad de la Teoría de Respuesta al Ítem el que, en los análisis que se derivan de ella, cada ítem de una prueba es en cierta forma

independiente del resto de ítems de la misma prueba, ofreciendo una medida de la dificultad “intrínseca” del ítem al margen de las poblaciones evaluadas. Por supuesto, para llegar a este punto se requiere una serie de procesos de pilotaje y ajuste de los ítems hasta llegar a su forma final, pero dicho esfuerzo vale la pena en la medida que no se cae en la seria limitación de la Teoría Clásica de los Test de restringir la dificultad de los ítems a la muestra particular y a la prueba en la que hayan sido testeados (es decir, si los ítems eran testeados con otros individuos o junto con ítems distintos en la prueba, su funcionamiento estadístico y su dificultad variarían). Por el contrario, un análisis siguiendo la Teoría de Respuesta al Ítem (por ejemplo, como el del modelo Rasch, que sigue la Oficina de Medición de la Calidad de los Aprendizajes²) suele reportar una escala donde se puede apreciar a todos los ítems ordenados por dificultad creciente. También entrega una lista de ítems que rompen la lógica de dicha escala de dificultad, lo que los vuelve pasibles de ser analizados y ajustados (esta era una de las tareas del Equipo de Evaluación, en el que se llevó a cabo la experiencia profesional).

3.1.4. Validez y confiabilidad de las pruebas

En su Marco de Fundamentación de las Pruebas de Evaluación Censal de Estudiantes (2016), la Oficina de Medición de la Calidad de los Aprendizajes concibe la validez como “un concepto unitario que alude al grado en que la evidencia empírica y la teoría apoyan la interpretación y el uso de las puntuaciones que son resultado de la aplicación de una prueba” (p.17), deslindando así de las concepciones tradicionales que entendían la validez como una característica únicamente del instrumento. De hecho, la validez se va forjando desde la definición misma de los constructos y debe mantenerse hasta el momento en que se analizan e infieren conclusiones a partir de los resultados obtenidos.

Por su parte, el mismo documento define la confiabilidad como “la consistencia de los resultados de una prueba a través de múltiples aplicaciones independientes” (p.18) y tiene estrecha relación con la validez. Cabe señalar que, debido a que las muestras seleccionadas para aplicar las pruebas de la Oficina de Medición de la Calidad de los Aprendizajes eran muy grandes, lo cual traía además una serie de complicaciones logísticas, se optaba por dejar

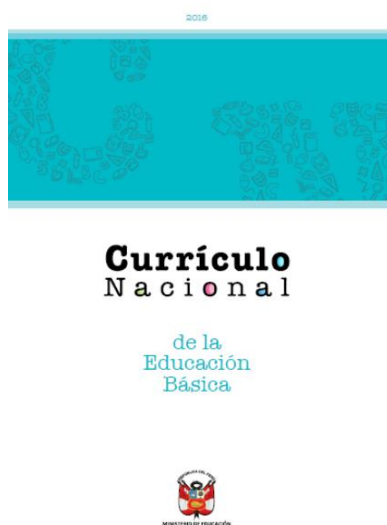
² Basándonos en el Marco de Fundamentación de las Pruebas de la Evaluación Censal de Estudiantes (2016), el modelo Rasch “parte del supuesto de que los resultados de la interacción entre personas e ítems no pueden estar totalmente predeterminados, sino que estos siempre implican un elemento de impredecibilidad”, lo que conlleva a que “en términos probabilísticos, mientras haya mayor habilidad, habrá mayor probabilidad de responder correctamente un ítem; y mientras más difícil sea un ítem, habrá menos probabilidad de acertarlo. Para ello, se establece un modelo matemático de tipo probabilístico que vincula la habilidad o rasgo latente de una persona con la probabilidad de respuesta correcta a un ítem” (p.71).

de lado métodos clásicos como el test-retest (es prácticamente imposible acceder a la misma muestra sin variaciones) o las formas paralelas (haría excesivamente grande el proceso de producción y pilotaje de ítems, considerando la temporalidad corta en la que este proceso debe desarrollarse). En lugar de ellos, la Oficina empleaba y aún emplea el modelo Rasch como método indirecto de consistencia interna, el cual calcula “la confiabilidad de las puntuaciones derivadas de un instrumento de evaluación” (p.18).

3.1.5. Constructos evaluados

Toda prueba psicológica debe definir, como primer paso, el o los constructos a evaluar, entendidos estos como el dominio o conjunto de rasgos observables a ser medidos por dicho instrumento. Las pruebas de la Oficina de Medición de la Calidad de los Aprendizajes, dado el carácter de ente evaluador perteneciente al Ministerio de Educación que tiene esta institución, toman como constructos a evaluar aquellos definidos en el Currículo Nacional vigente. En tal sentido, los constructos que se considera en las pruebas son las competencias que el Currículo señala como los aprendizajes que la escuela debe asegurar en los estudiantes (en el área curricular de Comunicación, las competencias consignadas son las de Comprensión lectora, Producción escrita y Comunicación oral). Dicho documento define competencia como “la facultad que tiene una persona de combinar un conjunto de capacidades a fin de lograr un propósito específico en una situación determinada, actuando de manera pertinente y con sentido ético”.

Figura 3. Currículo Nacional de Educación Básica (2016)



Las pruebas de la Oficina de Medición de la Calidad de los Aprendizajes buscan evaluar estas competencias reflejando sus principales componentes. Cuando este objetivo no se logra

del todo (por ejemplo, en el caso de la competencia de Comprensión lectora, las pruebas de la Oficina de Medición de la Calidad de los Aprendizajes que únicamente contemplan preguntas de opción múltiple se abstienen de evaluar la capacidad de Reflexión y evaluación de textos, una de las tres capacidades que conforman aquella competencia) dicha parcialidad es señalada de manera explícita en los marcos de evaluación de las pruebas. Cabe indicar que la forma como la Oficina de Medición de la Calidad de los Aprendizajes conceptualiza los constructos a evaluar, si bien se ampara en las competencias definidas por el Currículo Nacional de Educación Básica, también se ha visto enriquecida por las definiciones de experiencias internacionales de evaluación estandarizada en las que el Perú ha participado (las pruebas PISA o las del Laboratorio de UNESCO – LLECE).

3.2. ACCIONES, METODOLOGÍAS Y PROCEDIMIENTOS

El ingreso a la Oficina de Medición de la Calidad de los Aprendizajes pudo darse gracias a los conocimientos psicológicos ya adquiridos, los cuales se vieron ampliados y profundizados gracias a la experiencia adquirida en esta institución en relación con el proceso de aplicación de pruebas estandarizadas (principalmente de comprensión lectora), desde su construcción hasta el análisis de resultados. Dicho proceso implicó realizar las siguientes acciones y procedimientos.

3.2.1. Revisión y alineación de las pruebas con el documento curricular vigente

La Oficina de Medición de la Calidad de los Aprendizajes, en tanto unidad evaluativa del Ministerio de Educación, requiere coordinar y recibir aprobación de las unidades oficiales encargadas de la gestión curricular en dicho Ministerio (principalmente, la Dirección General de Educación Básica Regular), a fin de garantizar que los evaluados en las pruebas evalúen lo que el currículo oficial manda y lo que se supone que está siendo implementado en las aulas (al menos en el caso de las pruebas de carácter nacional, no así en el caso de las pruebas internacionales). Este proceso no está exento de complejidades, puesto que, especialmente en la época en la que se dio la presente experiencia laboral, el currículo para entonces vigente (Diseño Curricular Nacional 2009) tenía algunas deficiencias en la descripción de algunas competencias (como era el caso de Comprensión lectora y Producción escrita), por lo que de alguna forma se buscaba un equilibrio entre el respeto a la normativa vigente y la posibilidad de innovar o al menos corregir algunas deficiencias o

reduccionismos en la manera de concebir las competencias. Por ejemplo, el Diseño Curricular Nacional 2009 no contemplaba la variedad textual o de procesos lectores que la Oficina de Medición de la Calidad de los Aprendizajes, en parte debido a su contacto con experiencias internacionales de evaluación, sí intentaba introducir en sus pruebas a fin de dar cuenta de la comprensión lectora en un sentido amplio. En el caso de la Evaluación Censal de Estudiantes (ECE), iniciada el año 2007, luego de sus dos o tres ediciones anuales iniciales, se pudo llegar a un acuerdo estable y casi automático a nivel de todo el Ministerio en relación con lo que debía ser evaluado en estas pruebas. Estos procesos de “negociación técnica” eran llevados a cabo, en representación de la Oficina de Medición de la Calidad de los Aprendizajes, por los especialistas del Equipo de Evaluación.

3.2.2. Ítems de la prueba

La elaboración de ítems es una parte fundamental en el diseño de pruebas estandarizadas. Permite, en tanto cumpla con ciertos requerimientos técnicos, crear instrumentos confiables y pertinentes. Estos, a su vez, sirven para recoger evidencia específica sobre el nivel del logro de aprendizaje de los estudiantes. Por ello, en lo que respecta a las pruebas estandarizadas que permiten evaluar la competencia lectora de los estudiantes, se deben atender de modo integral cada una de las tres etapas de su preparación: la selección de textos relevantes para los grados evaluados, la edición de los mismos y la elaboración de ítems propiamente dichos.

Se denomina ítem a la unidad de medida con un estímulo y una forma prescriptiva de responder, que pretende elicitarse una respuesta en un sujeto a partir de la cual puede inferirse el desempeño de ese sujeto respecto de algún constructo psicológico (Osterlind, 1998). En el trabajo realizando elaborando pruebas estandarizadas de comprensión lectora en la Oficina de Medición de la Calidad de los Aprendizajes, los formatos de ítem más usuales eran los de construcción o producción de respuesta abierta, de selección o elección múltiple y los de verdadero-falso (Aiken, 2003; Osterlind, 1998).

Los ítems de construcción de respuesta demandan al estudiante que proporcione, desarrolle o cree una respuesta particular. Están formados por un enunciado, un espacio para responder, y en algunos casos van antecidos por un estímulo. Este consiste en una situación contextualizada o un texto (en el caso de Comprensión lectora) que le da un marco tanto al ítem como a la respuesta que se espera. También pueden incluir instrucciones o

especificaciones para responder. Además, este tipo de ítem demanda que se elaboren uno o más criterios de codificación, junto con ejemplos de respuestas que ilustren los niveles del criterio para poder valorar las respuestas de los estudiantes (Osterlind, 1998).

Figura 4. Ejemplo de ítem de respuesta abierta

Graffiti	
<p>Estoy indignada porque esta es la cuarta vez que han tenido que limpiar y volver a pintar la pared del colegio para quitar las pintadas. La creatividad es admirable, pero la gente debería encontrar otras formas de expresarse que no suponga gastos extra para la sociedad.</p> <p>¿Por qué nos creáis esa mala reputación a los jóvenes pintando en lugares en los que está prohibido? Los artistas profesionales no cuelgan sus cuadros en las calles, ¿verdad? Lo que hacen es buscar quien les financie y hacerse famosos a través de exposiciones que sí son legales.</p> <p>En mi opinión, edificios, vallas y bancos de los parques son obras de arte en sí mismos. Es realmente patético estropear su arquitectura con graffiti y lo que es más, el método que se utiliza para ello destruye la capa de ozono. Realmente no puedo entender por qué estos artistas delincuentes se enfadan cuando les quitan sus "obras de arte" de las paredes una y otra vez.</p> <p style="text-align: right;">Olga</p>	<p>Hay gustos para todo. Nuestra sociedad está invadida por la comunicación y la publicidad. Logotipos de empresas, nombres de tiendas. Gran número de carteles ilegales pegados por las calles. ¿Es esto aceptable? Sí, en general sí. ¿Son aceptables las pintadas? Algunos dirían que sí y otros que no.</p> <p>¿Quién paga el precio de las pintadas? ¿Quién paga al final la publicidad? Exacto. El consumidor.</p> <p>¿Acaso los que instalan las vallas publicitarias te han pedido permiso? No. ¿Y los que pintan los graffiti si tendrían que hacerlo? ¿No es todo una cuestión de comunicación, tu propio nombre, el nombre de las pandillas callejeras y las vallas publicitarias de las calles?</p> <p>Piensa en la ropa de rayas y cuadros que apareció hace algunos años en las tiendas y en la ropa de esquí. El estampado y los colores los habían copiado directamente de las floridas pintadas que llenaban los muros de cemento. Es bastante chocante que aceptemos y admiremos estos estampados y colores y que, en cambio, ese mismo estilo en graffiti nos parezca horroroso.</p> <p>Corren tiempos difíciles para el arte.</p> <p style="text-align: right;">Sofía</p>
<p>Pregunta 12:</p> <p>¿Por qué hace referencia Sofía a la publicidad?</p> <p>-----</p> <p>-----</p>	

Por su parte, el formato de selección múltiple (el más usual) presenta en su estructura un enunciado, las alternativas y un estímulo. El enunciado es una pregunta directa o una oración que debe completarse. Las alternativas son las opciones de respuesta entre las que el estudiante debe identificar la única que es correcta. Esto implica que dos o tres alternativas funcionan como distractores. El estímulo puede ser un texto o una situación específica en la que se plantea la tarea a resolver (Aiken, 2003; Osterlind, 1998).

Figura 5. Ejemplo de ítem de opción múltiple



El eucalipto es un árbol muy conocido que crece en todo el mundo. Se puede adaptar a diferentes climas, ya sean fríos, cálidos, secos o húmedos.

Es un árbol muy alto. Puede llegar a medir más de 100 metros. Tiene muchas ramas y hojas que dan una sombra agradable. A la gente le gusta esto y lo siembra en los parques.

Las hojas del eucalipto son alargadas y angostas. Además, si frota algunas hojas en tu mano podrás sentir un olor muy agradable, pues son fragantes. Este olor es bueno para curar la tos.

La madera del eucalipto sirve para hacer muebles resistentes porque es de muy buena calidad. Pero también puede arder con facilidad. Muchas veces, se producen incendios en los bosques donde hay eucaliptos.

Según el texto, ¿dónde crece el eucalipto?

- a Solo en climas secos.
- b En todo el mundo.
- c Solo en climas fríos.

Por último, el formato de verdadero-falso es un tipo de ítem de selección múltiple. En él, se presenta enunciados sobre un mismo contenido para que el estudiante determine la verdad o falsedad de las aseveraciones (Aiken, 2003; Osterlind, 1998). Este tipo de ítem ha sido utilizado solo en casos excepcionales en las pruebas de la Oficina de Medición de la Calidad de los Aprendizajes.

3.2.3. Codificación de respuestas

Para el caso de los ítems de construcción de respuesta abierta, se requiere llevar a cabo un proceso de codificación consistente y confiable de dichas respuestas. Dada la cantidad de respuestas que se reciben (en el caso de una prueba muestral, el promedio de individuos evaluados oscila alrededor de 2500-4000), se hace necesario contratar, capacitar y monitorear a codificadores que realicen esta labor. Estas funciones también fueron parte de mi experiencia en la Oficina de Medición de la Calidad de los Aprendizajes.

Para el caso de la capacitación de codificadores se utiliza material práctico que recrea el trabajo de codificación real que las personas convocadas y seleccionadas (la mayoría de ellos, docentes o psicólogos) habrán de realizar. Se les entrega unos manuales con los códigos por cada tipo de respuesta, así como ejemplos de respuestas de estudiantes (inventadas o reales) para que asignen los códigos que crean convenientes. Ese trabajo individual es revisado por los supervisores de la codificación (los especialistas del Equipo de Evaluación), además de socializado entre el grupo de codificadores para discutir los casos límite y aclarar el sentido de los códigos del manual. De esta manera, se garantiza que la codificación de las respuestas de los estudiantes se realice con criterios uniformes independientemente de quién sea el codificador.

El proceso de codificación propiamente dicho se realiza con los codificadores ya capacitados. Cada uno de ellos recibe una cantidad determinada de pruebas por día; además, tiene el encargo de revisar la codificación diaria realizada por uno de sus compañeros (revisa una muestra o el total de su cuota, según lo indique el coordinador de la capacitación). Estas parejas de revisión varían día a día: es otro mecanismo para optimizar la consistencia y uniformidad entre codificadores. Finalmente, una submuestra del total de pruebas es asignada a un proceso denominado “codificación múltiple”, en el que todos los codificadores codifican dicho set de pruebas. Es a partir de este proceso que, con ayuda del Equipo de Análisis y Factores Asociado, se analizan los valores de confiabilidad del proceso de codificación, en otras palabras, si es que los codificadores están alineados o no. En caso se detecten codificadores muy alejados de la media del grupo, sea porque son muy estrictos o muy permisivos a la hora de asignar los códigos a las respuestas, se les brinda en paralelo reforzamiento sobre el manual de códigos y su aplicación, entre otras estrategias de remediación.

Figura 6. Ejemplo de códigos de un ítem de respuesta abierta

Pregunta 2:

¿Cuál es la fecha de comienzo del gráfico en el Cuadro 1?

Sub-escala: Recuperar información
Dificultad: 540
Aciertos: España 32,8%;
OCDE 50,4%

1 punto: Respuestas que consignan 11.000 a.C. (o una aproximación entre 10.500 y 12.000) indicando que el alumno ha extrapolado a partir de la escala. Por ejemplo:

- 11.000.
- 11.000 a.C.
- 10.500 a.C.
- Justo antes de 10.000 a.C.
- Alrededor de 12.000.
- Alrededor de 11.000 a.C.

0 puntos: Otras respuestas, incluyendo flechas que apuntan al punto de comienzo del gráfico. Por ejemplo:

- 10.000 a.C. [No ha extrapolado a partir de la escala]
- 20.000 a.C.
- 8.000 a.C. [Se ha equivocado de figura]
- ~~11.000 a.C.~~ 4.000 a.C. [Ignórense las respuestas tachadas]
- 0.

3.2.4. Puntos de corte

En el caso de pruebas basadas en criterios, como las elaboradas por la Oficina de Medición de la Calidad de los Aprendizajes, el objetivo es el nivel de desempeño que presentan los estudiantes en el momento en que dichas pruebas son aplicadas, lo cual implica un proceso de ubicación de los estudiantes en determinadas categorías justamente a partir del rendimiento demostrado. Para ello, el Equipo de Evaluación, una vez recibida la información de las escalas de dificultad de los ítems, debe establecer “puntos de corte” en esa escala, lo que permitirá obtener los “niveles de desempeño” a los que serán asignados los estudiantes según su rendimiento en las pruebas. Estos puntos de corte funcionan como estándares o “puntos en la escala de las puntuaciones del test que sirve para clasificar a los sujetos en [...] categorías que reflejan diferentes niveles de competencia o conducta, con relación a un dominio” (Martínez Arias, 2005).

Existen diversas metodologías para realizar puntos de corte de manera adecuada. En el caso de la Oficina de Medición de la Calidad de los Aprendizajes; el método empleado es el de Angoff, que consiste en que “los jueces de la muestra de expertos examinen cada elemento del test y estimen la probabilidad de que una persona mínimamente competente responda a

cada ítem correctamente. [...] Los jueces piensan en términos de un amplio número de individuos mínimamente competentes y estiman qué proporción de este grupo respondería a cada elemento correctamente”. En el caso de la Oficina de Medición de la Calidad de los Aprendizajes, para realizar esta metodología, los integrantes del Equipo de Evaluación, junto con algunos especialistas externos invitados, asumían el rol de jueces y se reunían para decidir en qué punto de la escala de ítems ordenados por dificultad decidía establecer los puntos de corte. Luego de esta primera ronda, se socializaba los resultados y se pasaba a una segunda ronda, en caso algún juez decidiera cambiar su(s) punto(s) de corte. Se pasaba a una segunda y tercera socialización de resultados, y, de persistir algunas diferencias, se tomaban algunos acuerdos por mayoría o por decisión calificada del Equipo de Evaluación. En el caso de que, en años siguientes, se aplicara la misma prueba o una versión equiparable de la misma (como en el caso de la Evaluación Censal de Estudiantes), estos puntos de corte eran mantenidos a lo largo de dichos años justamente para permitir la comparabilidad de los resultados entre años.

Figura 7. Niveles de desempeño de la prueba ECE (Evaluación Censal de Estudiantes) de segundo grado de primaria en Comprensión lectora

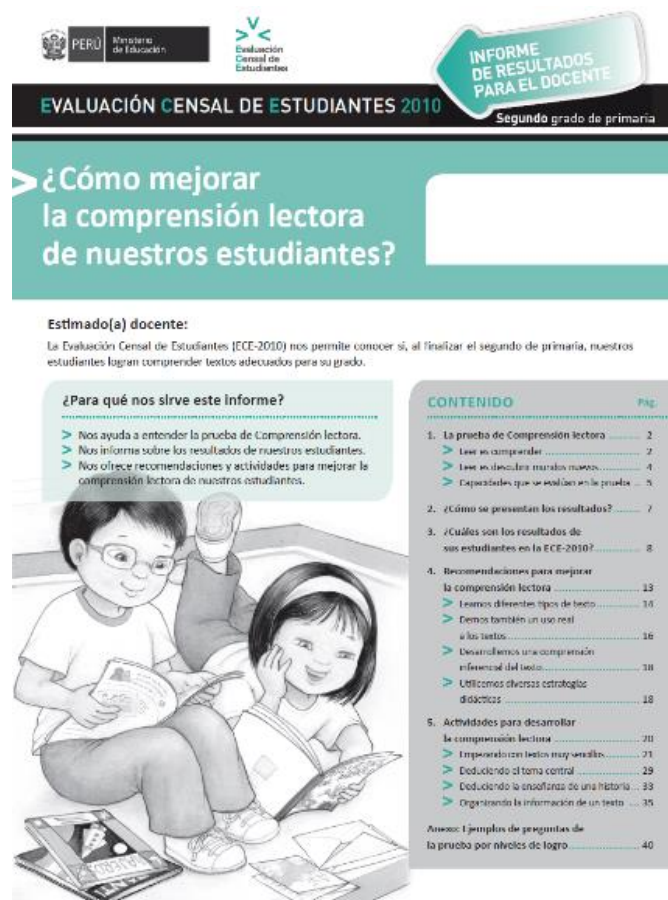


3.2.5. Reporte de resultados

Un proceso de evaluación estandarizada en el campo de la educación tiene sentido en la medida que, por un lado, es un proceso válido y confiable, y, por otro, la información que provee termina siendo de utilidad para los potenciales usuarios. En relación con lo último, la Oficina de Medición de la Calidad de los Aprendizajes se encarga de reportar los resultados de dos maneras (Marco de Fundamentación de las Pruebas de Evaluación Censal de Estudiantes, 2016):

- Por niveles de logro: Se indica qué porcentaje de los estudiantes evaluados se ubicó en cada uno de los niveles de logro establecidos a partir del proceso de puntos de corte anteriormente descrito.
- Por medida promedio: Se indica el puntaje promedio de los estudiantes de una escuela, provincia, región o del país en su conjunto. Esta manera de reportar resultados facilita mostrar las tendencias a lo largo del tiempo.

Figura 8. Portada de reporte de resultados de la prueba ECE (Evaluación Censal de Estudiantes) de segundo grado de primaria dirigido a docentes



CAPITULO IV: PRINCIPALES CONTRIBUCIONES

- La experiencia participando en la elaboración de pruebas estandarizadas de comprensión lectora en la Oficina de Medición de la Calidad de los Aprendizajes significó un importante aprendizaje y profundización en la realización de los pasos para el diseño de instrumentos válidos y la utilización de la información psicométrica para la optimización de los mismos.
- En cuanto a la definición de los constructos a evaluar, se consideraron las competencias consignadas en el Currículo Nacional de Educación Básica. En algunos casos, estas competencias pudieron ser evaluadas en su totalidad y en algunos casos solo parcialmente.
- En cuanto al diseño de las pruebas, se elaboraron diversos tipos de ítems, dando especial prioridad a los de opción múltiple y los de respuesta abierta. Estos últimos requirieron de la implementación de procesos de codificación de respuestas, a cargo de correctores entrenados.
- En cuanto al ajuste de las pruebas, se realizaron aplicaciones piloto y se contó con reportes psicométricos con indicadores estadísticos por cada ítem como orientador para la toma de decisiones en cuanto a mantener, modificar o descartar algún ítem de la prueba. En este caso, además, se produjo un aprendizaje referido a la interpretación de dichos reportes.
- En el caso de pruebas estandarizadas internacionales producidas en otro idioma, se realizó la traducción al castellano de las mismas, manteniendo la dificultad original de los ítems.
- Finalmente, una vez obtenidos los resultados de los estudiantes evaluados a través de las pruebas, se elaboró reportes dirigidos a público diverso en los que se comunicaban dichos resultados y se brindaban recomendaciones para la mejora de los aprendizajes de los estudiantes.

RECOMENDACIONES

- Los constructos a evaluar no deberían ser únicamente los del Currículo Nacional vigente (las competencias), sino que sería positivo que la Oficina de Medición de la Calidad de los Aprendizajes también incluya en sus pruebas la medición de variables favorecedoras

del aprendizaje (ampliamente estudiadas por la Psicología), como la motivación, la autoestima, el autoconcepto, entre otros, las cuales, si bien no son competencias, enriquecerían la información recogida y reportada por la sobre la realidad de los estudiantes peruanos.

- En cuanto al diseño de las pruebas, se podrían incluir nuevos tipos de ítems asociados a modos de aplicación de las pruebas en entornos virtuales. Por ejemplo, la aplicación de pruebas en computadora permitiría, en el caso de la comprensión lectora, que el estudiante extraiga manualmente la información del texto que se le solicite, con lo cual se reduce la posibilidad de problemas a la hora de expresar por escrito una respuesta.
- En cuanto al ajuste de las pruebas, sería conveniente también publicar los reportes de las aplicaciones piloto para que el público interesado puede rastrear con mayor detalle todo el proceso de afinamiento que han pasado los ítems considerados en la versión final de la prueba, además que sería un insumo importante para alfabetizar en temas psicométricos a más especialistas en este campo de la psicología.
- En cuanto a la traducción de pruebas de idiomas extranjeros al castellano, convendría realizar algunos procesos de verificación de que dichas traducciones realmente son compatibles con los distintos dialectos hablados en el país.
- En cuanto a los reportes de resultados producidos, estos podrían cada vez más ser digitalizados y habilitar un acceso interactivo a los mismos, de tal manera que cualquier usuario pueda personalizar sus búsquedas para sacarle mayor provecho a toda la información producida.

REFERENCIAS BIBLIOGRÁFICAS

Aiken, Lewis (2003). Tests psicológicos y evaluación. México: Pearson.

Martínez Arias, Rosario (2005). Psicometría: Teoría de los tests psicológicos y educativos. Madrid: Editorial Síntesis.

Ministerio de Educación (2016). Currículo Nacional de la Educación Básica.

Ministerio de Educación (2016). Marco de Fundamentación de las Pruebas de la Evaluación Censal de Estudiantes.

Osterlind, Steven (1998). What is constructing test items? Dordrecht: Springer.

Popham, James (1978). Criterion-Referenced Measurement. New Jersey: Prentice-Hall.

Silva Moreno, Fernando (Ed.) (2003). Evaluación psicológica en niños y adolescentes. Madrid: Editorial Síntesis.

