

### Capítulo III.

## Administración y resultados de las pruebas

La administración de las pruebas en el ámbito nacional, en 60 sedes del país, fue convocada para el día 1 de marzo del año 2002. Se esperaba que en esa fecha sean examinados todos los concursantes inscritos que habían obtenido en la primera fase, constituida por la evaluación de los currículos profesionales y la entrevista personal, al menos un puntaje mínimo para tener derecho a pasar a la fase del examen escrito. Sin embargo, sólo se presentaron a rendir las pruebas en esa fecha 57.007 concursantes que tuvieron acceso a los locales programados y contaron con los cuadernillos de preguntas y las hojas de respuestas necesarias para rendir los exámenes.

Empero, como fue públicamente conocido después, 38.212 concursantes no pudieron ser evaluados en 14 sedes (Chachapoyas, Huaraz, Chimbote, Pomabamba, Abancay, Andahuaylas, Cusco, Huánuco, Huancayo, Trujillo, Chiclayo, Iquitos, Mariscal Ramón Castilla y Sullana) debido a los actos de oposición a la administración de las pruebas de suficiencia profesional que condujo el SUTEP en todo el país. Las medidas de fuerza de este sindicato tuvieron un rango de variación que incluyó desde la toma de iglesias y huelgas de hambre de algunos de sus miembros hasta la toma de locales escolares y la captura y quema pública de los cuadernillos de preguntas y de las hojas de respuestas. Estos actos en un número importante de casos, en nuestra opinión, contaron con la anuencia, por acción o por omisión, de las autoridades locales que presumiblemente percibían con disgusto la administración de las

[57]

pruebas debido a que la calificación anónima y automática de las mismas les privaba de la posibilidad de seleccionar, según sus preferencias, a los ganadores del concurso aprovechando el conocimiento que les brindaron las relaciones cara a cara que imperaron durante la primera fase del concurso.

Con el propósito de superar esta situación, el Ministro de Educación programó una segunda fecha para el día 9 de marzo del año 2002 y solicitó al equipo de trabajo encargado de las pruebas que se mantenga por una semana más en situación de incomunicación a fin de preparar un nuevo juego de pruebas, equivalentes en lo posible en contenido y grado de dificultad a las que se habían administrado el día 1 de marzo. De este modo se elaboró nuevamente pruebas especializadas en las 14 menciones de formación docente que incluía entonces nuestro sistema educativo (ver cifras en las tablas n.º 1 y n.º 2). Como medida adicional, para garantizar que las autoridades locales presten apoyo decidido a la administración de las pruebas, el Ministro de Educación suspendió en sus funciones a los directores regionales y a los directores zonales de las sedes en conflicto, encargando temporalmente estas funciones a personal de su máxima confianza.

### **El experimento no previsto**

La situación antes descrita dio oportunidad no planificada para someter a control experimental la calidad de las pruebas que se administraron. Ocurrió que en la primera fecha, en la ciudad del Cusco, rindieron sus pruebas satisfactoriamente 2.234 docentes pero se frustró el examen para 5.103 concursantes que por los actos de violencia no pudieron concluir con la entrega de sus hojas de respuesta. Por una especial coyuntura política el Ministro de Educación accedió ante el gremio de docentes del Cusco que en la segunda fecha (9 de marzo) fueran examinados no sólo los 5.103 que no habían dado las pruebas aún sino que también se les tome examen, otorgándoles una segunda oportunidad, a los 2.234 concursantes que habían dado su prueba satisfactoriamente en la primera fecha. Este pedido del gremio docente, probablemente, se debió a que los examinados en la primera fecha percibieron que las pruebas administradas tenían un alto índice de dificultad en relación con sus conocimientos especializados. Consideraron, presumi-

blemente, que con una semana adicional podrían mejorar su preparación o que, tal vez, dado el brevísimo lapso con el que el Ministerio de Educación contaba para la segunda administración, recurriría a reimprimir las mismas pruebas.

Sin embargo, lo que ocurrió en los hechos fue que el equipo de trabajo elaboró y puso a disposición del Ministerio de Educación un nuevo juego completo de pruebas. De esta manera se configuró la situación experimental porque el día 9 de marzo del 2002 dieron examen inequívocamente identificados 2.234 docentes que se sometieron voluntariamente a las dos versiones de las pruebas, las mismas que eran materialmente distintas pero, a juicio del equipo constructor de las pruebas, teóricamente equivalentes en tanto que las preguntas habían sido seleccionadas observando los mismo criterios en lo referente a su contenido cognoscitivo, grado de dificultad y poder discriminativo.

Durante el procesamiento de datos se calculó el promedio de rendimiento en la primera y segunda versión de las pruebas de los 2.234 docentes examinados. Sobre un máximo de 40 puntos obtuvieron 12,603 en la primera fecha y 12,013 en la segunda. La diferencia entre los promedios de las dos administraciones fue sólo de 59 centésimos, lo que significa que, prácticamente, el rendimiento académico de los 2.234 docentes de la región Cusco, en ambas versiones de las pruebas, fue estadísticamente el mismo, resultado que autoriza a pensar que las 14 pruebas de la primera fecha fueron casi equivalentes en grado de dificultad a las 14 de la segunda fecha. Esta estabilidad en la medición se muestra en la tabla n.º 16. Se condice fuertemente con los altos índices de confiabilidad y bajos índices de error estándar que se registran en la tabla n.º 7.

### **Distribución de la población**

Considerando por separado cada una de las pruebas, el mayor número de concursantes, 34.631, se sometió a la prueba correspondiente a educación primaria. Si asumimos como especialidades típicas de la educación secundaria-matemática (D), comunicación (E), ciencias sociales (F), ciencias naturales (G), educación artística (H) y lengua extranjera (I) encontramos que concursaron al menos 36.149 docentes secundarios. Las especialidades de ciencias naturales y de ciencias sociales fueron las más numerosas en el nivel secundario

con 9.498 y 9.205 concursantes, respectivamente. Llama la atención el muy reducido número de profesores de lengua extranjera que se presentaron al concurso, 809 en total. La prueba de lengua extranjera, para fines prácticos de idioma inglés, fue solicitada por un número muy reducido de docentes si tenemos en cuenta que sus servicios son necesarios tanto en primaria como en secundaria y el Ministerio de Educación registra aproximadamente 60 mil centros educativos públicos. Asimismo, es destacable que el número de concursantes del área de matemática, que totaliza 7.669 docentes, es inferior al número de concursantes de ciencias naturales y de ciencias sociales, siendo que ésta es una especialidad que debería tener una oferta de servicios mayor si se considera su carácter formativo y su valor estratégico en todas las comparaciones internacionales relevantes sobre la eficiencia del sistema educativo. El mismo comentario merece la especialidad de comunicación.

Desde el punto de vista territorial la población estuvo dividida en 60 sedes que no corresponden exactamente a 60 ciudades debido a que siete distritos de Lima Metropolitana fueron contabilizados como sedes independientes (ver la tabla n.º 3). Sin embargo, si sumamos los concursantes que rindieron examen en dichos distritos y en la jurisdicción directa de la Dirección de Educación de Lima, constatamos que en la ciudad capital se sometieron a las pruebas 15.136 docentes que representaron el 15,9% de la población examinada. La ciudad de Puno, con 8.831 docentes, ocupó el segundo lugar. El tercer lugar le correspondió al Cusco con 7.337 docentes. Luego siguieron Trujillo, Cajamarca, Arequipa y Huancayo con 5.806, 5.070, 4.976 y 4.610, respectivamente.

Número de concursantes en siete principales ciudades

<i>Ciudades</i>	<i>N.º de docentes concursantes</i>	<i>Porcentaje</i>
Lima	15.136	15,9
Puno	8.831	9,2
Cusco	7.337	7,7
Trujillo	5.806	6,1
Cajamarca	5.070	5,3
Arequipa	4.976	5,2
Huancayo	4.610	4,8
Total	51.766	54,3%

Esto significa que en las siete ciudades más importantes del país se sometieron al examen 51.766 concursantes que representaban el 54,3% de la población examinada (95.219). Asimismo, el 28% tenía el título de Licenciado en Educación otorgado por el sistema universitario, el 68% poseía el título de Profesor de Educación Secundaria, Primaria o Inicial expedido por los institutos superiores pedagógicos, el 1,92% declaró carecer de título pedagógico y el 1% se abstuvo de precisar su calificación profesional.

### El proceso de calificación y su avatar

La calificación de las prueba de suficiencia profesional se realizó siguiendo un procedimiento derivado de las normas metodológicas que la bibliografía especializada establece para determinar los puntajes asignables a los instrumentos psicométricos de opción múltiple. En este sentido y aplicando el reglamento (D.S. 065-2000-ED), que establecía una nota máxima de 40 puntos, se asignó 0,80 por cada acierto, el descuento de 0,20 por cada error y 0 puntos por la pregunta no abordada. El descuento por los errores, como es conocido, se justifica para descontar la probabilidad que tiene un sujeto examinado de emitir una respuesta correcta por azar. En el caso particular de las pruebas de suficiencia profesional, debido a que cada pregunta tenía cinco opciones de respuesta, una correcta (clave) y cuatro distractores, la probabilidad de acertar por azar fue igual a  $\frac{1}{4}$  del valor de la pregunta.

Empero, la dificultad mayor para utilizar los resultados de las pruebas en la distribución de las 28.125 (ver la tabla n.º 5) mil vacantes ofertadas para nuevos nombramientos y para comunicar a la opinión pública un diagnóstico sin alteraciones de la situación académica del magisterio nacional surgió debido a que el artículo 15º del reglamento también establecía que para nombrar a un docente en una plaza éste debería necesariamente haber obtenido nota aprobatoria, lo que, en función de la situación descrita en el párrafo sobre «Puntajes mínimos a priori», significaba al menos 22 puntos que equivalen a 11 puntos en la escala vigesimal.

Esta condición reglamentaria creó al Ministro de Educación una dificultad política de magnitud debido a que solamente 2.837 concursantes habían obtenido 22 puntos o más y era voluntad del

gobierno entregar entre 15 mil y 18 mil vacantes de las 28.125 ofertadas a través de la publicación de resultados que mostraran la rectitud de la autoridad en el cumplimiento del reglamento; esto es, todos los nombrados, que deberían ser una cifra cinco o seis veces mayor que la de los realmente aprobados en el examen, deberían figurar en las nóminas oficiales con 22 puntos o más. Como se comprende, era imposible, sin alterar los resultados, nombrar a entre 16% y 19% de la población con notas iguales o mayores a 22 puntos cuando menos de un 3% estaba en esta condición. La decisión política fue asignar vacantes a todos los concursantes que hayan obtenido en el ámbito de su sede un puntaje igual o mayor al promedio real de esa sede, añadiendo un factor de corrección al puntaje real que había obtenido cada postulante, lo que dio lugar a un número de aprobados que excedió al número de vacantes ofertados. Este factor de corrección para una sede  $S$  ( $1 = S = 60$ ) libremente elegida fue igual a la diferencia que existía entre 22 puntos y el promedio de dicha sede  $S$  ( $1 = S = 60$ ) en una determinada prueba  $P$  ( $1 = P = 14$ ).

En tanto que ninguno de los 840 promedios posibles (considerando 60 sedes y 14 pruebas) alcanzó un valor mayor de 22 puntos, con excepción de la sede 47 en la prueba A, el factor de corrección generalmente fue positivo y, naturalmente, en las sedes en las que los promedios fueron más bajos el factor de corrección tuvo un valor sensiblemente mayor (ver tablas n.ºs 10 y 15) se muestran los factores de corrección usados en cada sede para cada una de las pruebas.

Es claro que esta decisión política dio lugar a que miles de miembros del magisterio se sintieran aprobados y con más derecho de los que les correspondían por desconocimiento de los resultados reales de la medición practicada. Así estuvieron impedidos de cobrar conciencia de los muy acusados déficit académicos que afectan su formación profesionalmente. Sin embargo, quizá la consecuencia más grave fue que se impidió que la comunidad nacional y los sectores dirigentes conozcan un indicador que además de condecirse con otros elaborados por las comunidad internacional apunta hacia la identificación de las causas explicativas de los pobres rendimientos de nuestros escolares en las pruebas UNESCO (1998-2000) y pruebas PISA (2001). Empero, aunque el procedi-

miento elegido no fue técnicamente el mejor, debemos reconocer que hubo una justificación que sonó defendible en función de la razón práctica en el sentido de que era una medida de justicia el otorgar nombramientos a miles de profesores afectados por la inestabilidad laboral aun cuando adolezcan de insuficiencias académicas que, además, no son de responsabilidad personal sino sistémica. El procedimiento elegido era imparcial y equitativo en tanto que el criterio de referencia para los cálculos no era decidido por la autoridad sino por el rendimiento real de los concursantes en cada sede y en cada prueba. Asimismo, el factor de corrección era de mayor magnitud en las sedes con rendimientos más bajos, lo que compensaba el hecho de que estas mismas sedes eran afectadas severamente por sus altos índices de pobreza extrema.

#### **Validez y confiabilidad de las pruebas**

Una de las preocupaciones de instituciones como Transparencia y la Defensoría del Pueblo fue inquirir sobre la validez de las pruebas. Gravitó sobre esta inquietud el hecho de que en el Ministerio de Educación, en las facultades de Educación y en los institutos superiores pedagógicos es frecuente que funcionarios, profesores y alumnos hablen de «pruebas validadas». Es oportuno señalar que, hasta donde hemos podido indagar, la manera como se usa en dichos medios el concepto de pruebas validadas generalmente expresa un desconocimiento de los procedimientos metodológicos que se usan en psicometría para determinar los índices de confiabilidad y validez de las pruebas. En primer lugar, es importante aclarar que los conceptos de confiabilidad y validez en la bibliografía especializada están ligados fuertemente a la utilización de un procedimiento experimental en series de tiempo. Esto significa que en un país en el que no existe antecedentes registrados de administración de pruebas educacionales a poblaciones masivas resulta fuera de lugar pretender, como ocurrió en algunos casos, preguntar por los índices de validez y confiabilidad de la primera experiencia de administración masiva de pruebas nacionales especializadas a un universo de 95.219 personas, siendo que estos instrumentos fueron preparados, contra las normas convencionales, en menos de dos meses.

En efecto, para poder demandar los índices de confiabilidad y validez de los instrumentos psicométricos que se usan en el ámbito nacional se requiere de un trabajo experimental previo, el mismo que se realiza dentro de una institucionalidad académica y científica que, lamentablemente, no existe ni siquiera en ciernes en el Perú. Para nuestros fines será suficiente por ahora señalar que «confiabilidad de una prueba» se define por el grado de estabilidad con que la misma asigna puntajes a los sujetos examinados. Este grado puede ser estimado mediante el análisis de los resultados, obviamente, después de que la prueba ha sido administrada, aplicando procedimientos estadísticos como el propuesto por Kuder y Richarson que se resumen en la fórmula KR20, que hemos utilizado en este caso y cuyos resultados se encuentran detallados en la tabla n.º 7. Sin embargo, los valores obtenidos en una situación como la descrita en este informe interpretan, correctamente analizando su significado dentro de un conjunto amplio de indicadores, que también en este caso son conocidos en la fase posterior a la administración de las pruebas.

Ciertamente, el procedimiento KR20 fue el mejor que tuvimos a la mano considerando que no estaba a nuestro alcance, por razones políticas y administrativas, usar los métodos más adecuados que son los experimentales. Ello hubiera requerido que planifiquemos, con los reajustes del caso, la readministración por muestreo de cada una de las 28 pruebas en series de tiempo, lo que resultaba inviable en una situación de inestabilidad política en la que en dos años tuvimos tres ministros de Educación.

Respecto de la validez de las pruebas, es pertinente subrayar que este concepto psicométrico es menos simple que el concepto de confiabilidad. Usualmente se define la validez de una prueba como la propiedad que ésta tiene de medir aquello que pretende medir. Esta definición no entraña especiales dificultades cuando lo que pretendemos calcular son magnitudes tangibles y macroscópicas, como por ejemplo medir la estatura o el peso de un estudiante. Pero cuando se trata de medir capacidades no tangibles como la inteligencia de una persona o sus suficiencias para ejercer la docencia, entonces la dificultades epistemológicas, teóricas y metodológicas no son simples en tanto que sólo podemos limitarnos a medir las capacidades o potencialidades intangibles a partir de un conjunto

de indicadores observables que permiten deducir que la capacidad examinada es efectivamente poseída por los sujetos investigados. Esta solución no está libre de dificultades relevantes en tanto que la selección de los desempeños o tareas que revelan que existe una cierta capacidad o competencia en una persona y en un determinado grado puede hacerse utilizando diversos criterios que, inclusive, podrían ser incompatibles entre sí.

Para superar estas limitaciones, los investigadores L. J. Cronbach y P. E. Meehl en un famoso artículo titulado *Validez de los Constructos en los Test Psicológicos*, publicado en 1956, distinguieron cuatro categorías: validez predictiva, validez concurrente, validez de contenido y validez de constructos. Como se deduce de esta clasificación conceptual, para obtener valores para los indicadores de validez se requiere de un trabajo especializado de mayor complejidad que cuando se investiga la confiabilidad de un instrumento. Por ejemplo, la validez predictiva demanda una investigación de seguimiento en series de tiempo para determinar si los sujetos que el instrumento ha seleccionado como poseedores de una capacidad, efectivamente demuestran tenerla a través de sus desempeños observados durante un lapso previamente definido. Este trabajo, por su duración, sólo es posible si existe una institucionalidad que propicia la especialización y la dedicación a las investigaciones psicométricas. Lo mismo puede afirmarse de la validez concurrente que presupone la existencia de criterios que provienen de los registros de investigaciones realizadas con otras pruebas o con una base de datos de récords académicos y profesionales preexistentes.

La validez de contenido se define como la propiedad de un instrumento de recoger a través de sus preguntas, de manera equilibrada y representativa, todos los contenidos temáticos que los sujetos examinados deberían conocer. Este tipo de validez depende del juicio de los especialistas y de la estabilidad y claridad con que se definen las áreas materia de medición. El trabajo correspondiente puede hacerse satisfactoriamente en un gabinete si se cuenta con el equipo de trabajo idóneo y con la información documental y bibliográfica adecuada. En el caso de las pruebas de suficiencia profesional tuvimos que afrontar dos dificultades relevantes. La primera estuvo constituida por la acusada inestabilidad que afec-

ta a los currículos oficiales de formación magisterial y de educación básica que son una primera fuente de información para medir las competencias de los docentes para dirigir el aprendizaje. La situación se presentó difusa y desordenada debido a que en varias áreas había más de un currículo vigente y no estaba claro su ámbito de aplicación.

La segunda dificultad estuvo dada por el hecho de que los funcionarios del Ministerio de Educación que trabajan en materia curricular no son, normalmente, miembros de la comunidad científica y, en consecuencia, ocurre que los currículos oficiales están integrados en la última década por información científica, tecnológica y humanística desfasada, mutilada y lógicamente desarticulada, base sobre la cual es muy difícil construir pruebas de suficiencia profesional válidas. Por otro lado, los miembros de la comunidad científica convocados, matemáticos, biólogos, químicos, lingüistas, antropólogos, sociólogos, etc., se dedican a la investigación y a la enseñanza universitaria pero desconocían la documentación y la institucionalidad que rige la formación magisterial y los estándares de aprendizaje de los centros educativos. Para intentar superar las limitaciones antes descritas fue necesario realizar un trabajo de seminario integrando a especialistas en currículo del Ministerio de Educación y a miembros de la comunidad científica en sesiones continuas y en situación de internamiento, por 15 días, para elaborar pruebas cuyo contenido posea calidad científica y adecuación, en lo posible, a las características de los currículos vigentes.

De lo expuesto se deduce que fue materialmente imposible estimar valor alguno de validez predictiva o de validez concurrentes para las pruebas de suficiencia profesional. Cualquier esfuerzo en esa dirección requería de un trabajo de investigación experimental que no interesó al Ministerio de Educación agobiado por la coyuntura política y que, además, requería que se dedique el presupuesto al análisis causal de resultados impopulares que eran y siguen siendo incómodos para la dirigencias sindicales en sus demandas de incrementos salariales; ello debido a que la deficiente calidad de la formación profesional del magisterio, evidenciada por las pruebas, no justificaba aumentos indiscriminados, electoralmente atractivos, pero inadecuados para esti-

mular el esfuerzo personal de los docentes por mejorar su formación académica.

La incomodidad parecía aún mayor en las instituciones que se han beneficiado económicamente con la tercerización del servicio de capacitación docente por una década, a través del Plancad, las mismas que tienen responsabilidad directa junto con los cuadros directivos del Ministerio de Educación en la desorientación y debilidad académica de la que adolece el sistema de formación docente y el de capacitación a cargo de Dinfocad. Lo que nos ha sido posible por ahora, a través del análisis de gabinete, es emitir algunos juicios acerca de la validez de constructos de las pruebas en base al análisis combinado del índice de discriminación y del índice de dificultad de cada pregunta.

El índice de discriminación informa si un instrumento psicométrico tiene capacidad para diferenciar el grupo de alto rendimiento del grupo de bajo rendimiento. Los valores de este índice varían entre uno y menos uno. La bibliografía especializada considera que un valor positivo igual o mayor a 0,25 es suficiente para afirmar que la pregunta dada tiene un aceptable poder discriminativo. Como se indica en la tabla n.º 16, el valor promedio de los índices de discriminación de las pruebas administradas en la primera fecha (2 de marzo) es casi cuatro centésimos menor que el mínimo aceptable. Sin embargo, lo más importante en este caso es el valor promedio del índice de discriminación de cada prueba. Así tenemos que las pruebas B, C, E tuvieron un comportamiento muy por debajo del mínimo aceptado, lo que significa que en los dos primeros casos los instrumentos carecieron de poder discriminativo y en el tercer caso el instrumento tendió a discriminar negativamente, lo que se interpreta en función de que las preguntas más difíciles favorecieron más a los estudiantes de bajo rendimiento, posiblemente debido a su ambigüedad.

Sin embargo, la situación anterior mejoró en la segunda fecha (9 de marzo) en tanto que la prueba de ciencias naturales pasó de un valor negativo a un alto valor positivo (0,48277147), el mismo que resulta bastante satisfactorio en la medida que revela que esta prueba mejoró notablemente al disminuir su grado de dificultad y tuvo capacidad relevante para diferenciar a los estudiantes de bajo rendimiento de los estudiantes de alto rendimiento. Asimismo, en

esta segunda administración nueve pruebas (B, C, E, F, G, I, J, K, N) tuvieron índices de discriminación con un valor promedio superior al mínimo establecido por la bibliografía especializada. Toda las demás pruebas obtuvieron valores que, en términos estadísticos, equivalen al mínimo. Evidentemente, esta vez no hubo prueba con índice negativo.

El índice de dificultad de una prueba es un indicador del grado en que las preguntas se encuentran dentro de la capacidad de respuesta de los examinados ya sea por el grado de complejidad de los contenidos, procedimientos y mecanismos que se examinan o por la mayor o menor familiaridad de los usuarios con los temas que constituyen su contenido. Este índice varía entre 0 para las preguntas que respondieron todos los examinados y 1 para las preguntas que no fueron respondidas por sujeto alguno. La bibliografía considera que las preguntas que tienen valores entre 0,51 y 0,60 tienen un índice de dificultad *muy moderado* y las preguntas que asumen valores entre 0,61 y 0,70 un grado de dificultad *moderado*. La lectura de la tabla n.º 20 indica que el índice de dificultad promedio de las pruebas ha sido estadísticamente moderado y adecuado para seleccionar al menos al tercio superior de la población examinada. Es importante aclarar que hay una relación estadística entre el índice de dificultad y el índice de discriminación. Frecuentemente las preguntas con un grado de dificultad mayor que 0,8 carecen de valor discriminativo significativo.

Grados de dificultad de las preguntas  
considerando que las pruebas debieron seleccionar  
al 29,54% superior de la población

Intervalos	Grado de dificultad
0,51 - 0,60	Muy moderado
0,61 - 0,70	Moderado
0,71 - 0,80	Difícil
0,81 - 1,00	Muy difícil

Por otro parte, a fin de ubicar las dificultades que han afectado severamente la elaboración de las pruebas de suficiencia profesional en el contexto de la comunidad internacional, nos permitimos

señalar que en los Estados Unidos de Norteamérica y en Europa hay una cultura de respeto a las pruebas de evaluación, como instrumentos eficientes, imparciales y equitativos de medición de aprendizajes y reconocimiento de méritos, que se desarrolla conducida por una institucionalidad que tiene más de 50 años de antigüedad. Asimismo, en la región de América Latina y el Caribe hay muchos países en los que existen centros e institutos dedicados a la elaboración de pruebas educacionales de administración masiva hace alrededor de 40 años. Por ejemplo nuestros vecinos Brasil, Colombia y Chile tienen centros, institutos y departamentos universitarios dedicados a la investigación científica en materia de pruebas educacionales y a la producción de las mismas que, administradas en el ámbito nacional, evalúan los aprendizajes de los estudiantes de educación básica y de educación superior. Es el caso del Instituto Nacional de Evaluación y Pesquisas Anisio Teixeira de Brasilia, del Instituto Colombiano para el Fomento de la Educación Superior y del Ministerio de Educación de Chile que tiene unidades dedicadas a la producción anual de las pruebas PSU para la educación superior y de las pruebas SIMCE (Sistema Internacional de Medición de la Calidad de Educación) para educación básica.

Aquí cabe mencionar que la metodología para la elaboración, administración y calificación de pruebas de administración masiva en la región fue difundida y perfeccionada en lengua castellana por el *College Entrance Examination Board*, de Puerto Rico que inicio hace más de 40 años la elaboración y aplicación de la pruebas de actitud académica para seleccionar a los postulantes que solicitaban ingreso a las universidades.

### **Indicadores generales de las pruebas de suficiencia profesional: promedios**

Considerando la población de 95.219 concursantes evaluados y 28 pruebas de suficiencia profesional administradas, subdivididas en 14 tipos correspondiente a las 14 especialidades oficiales de formación magisterial, el promedio general de rendimiento obtenido para la población es de 11,9230 como figura en la tabla n.º 6, el mismo que especifica valores por género asignando 11,49 a los

varones y 12,19 a las damas. La diferencia entre estos dos últimos promedios es menor que 1 y poco significativa, aunque tiende a ser favorable a las damas. La desviación estándar de 5,24 para varones y 6,07 para las mujeres revela que el grupo de los varones es ligeramente más homogéneo que el de las damas.

El error estándar de cada una de las pruebas, tal como aparece especificado en la tabla n.º 7, es en todos los casos menor que 4, lo que significa que de acuerdo con la bibliografía especializada es satisfactorio. Esta característica se acentúa si consideramos que las pruebas del tipo E, en la primera fecha, tienen un error estándar menor que 2 y que se registran, en las dos fechas, nueve pruebas con un error estándar menor que 3. Asimismo, considerando que en los cálculos se ha asumido un máximo de 40 puntos, el déficit de la formación académica de los docentes es de aproximadamente el 70,02%, tomando como referencia la máxima calificación esperada; y si se toma como referencia minimalista la nota aprobatoria de 22 puntos, alcanza hasta un 46%.

Estos valores estadísticos en las fechas en las que se administraron las pruebas constituyeron una constatación dolorosa para quienes han tenido la responsabilidad de conducir, desde diferentes roles, el Ministerio de Educación del Perú durante los últimos 25 años. Una manera de evadir el malestar fue argumentar que se trataban de datos aislados proporcionados por pruebas que no habían sido previamente «validadas». Sin embargo, algunos meses después se hicieron públicos los resultados de las investigaciones evaluativas internacionales realizadas por la UNESCO entre 1998 y el año 2000 y las efectuadas por la OECD a través de la aplicación de las pruebas PISA en el Perú el año 2001, conociéndose tablas (*ranking*) con los resultados finales que ubican al sistema educativo peruano, en el nivel primario, en el último lugar, junto con Haití, en la región de América Latina y el Caribe y en el último lugar de un conjunto de 43 países representativos de los cinco continentes que participaron en las pruebas PISA diseñadas para el intervalo de 15-16 años de edad.

Evidentemente, la situación considerablemente precaria de los aprendizajes de los escolares peruanos mostrada por la comunidad internacional no podía entenderse independientemente de la calidad de la formación profesional de sus maestros. Los resulta-

dos lamentables que hemos obtenido con las pruebas de suficiencia profesional aportaron la descripción de la otra cara de la medalla porque guardan una relación de correspondencia simétrica con la ubicación extremadamente deficiente que la evaluación de UNESCO y OECD asigna a nuestro sistema educativo. En breve, los hechos han rebasado el poder de quienes administran el silencio y la desinformación y han dado evidencias sólidas de que las pruebas de suficiencia profesional, en su conjunto, han demostrado que la precariedad académica de los escolares peruanos es un fiel reflejo de la de sus profesores.

Desde otra perspectiva, los resultados dan una evidencia fáctica de los desequilibrios socioeconómicos que fracturan a la sociedad peruana. En la tabla n.º 9 se muestra que tomando como referente un valor máximo de 40 puntos, el promedio más alto de rendimiento lo obtuvo la Use 07 de San Borja con 14,82098 y el promedio más bajo lo tiene la subregión Requena con 7,57886 precedida por la subregión Sara Sara-Cora Cora con un promedio 8,09416. Cabe mencionar que según el mapa de la pobreza publicado por el gobierno peruano en los periódicos de Lima el 11 de febrero de 2005, Loreto tiene un 40,8% de pobreza extrema y la región Ayacucho un 41,8%, mientras que la región Lima figura con un 3,8% de extrema pobreza.

Lo sorprendente y en apariencia paradójico es que la subregión Loreto-Nauta haya alcanzado un promedio de 14,43261 ocupando el segundo lugar en el ámbito nacional y su ciudad vecina Iquitos haya ocupado el tercer lugar, también en el ámbito nacional, con 14,31124. Esta dos sedes amazónicas, que culturalmente forman solamente una, tienen puntajes superiores a todas las otras sedes de Lima y de las ciudades principales del Perú que son Arequipa, Trujillo y Cusco. Hay buenas razones para pensar que la posición destacada de Nauta e Iquitos no es casual porque la información proporcionada por la tabla n.º 19 muestra que Iquitos obtuvo los promedios máximos en el ámbito nacional en las pruebas de matemática, ciencias naturales, comunicación y en la prueba para los centros de educación ocupacional. Dentro de este conjunto, las áreas de matemática y comunicación, según la opinión de la bibliografía especializada y de las instituciones que elaboran pruebas internacionales, son representativas de la calidad del sistema

educativo en su conjunto. En consecuencia, sería de mucho provecho para el país realizar investigación educacional en las ciudades de Iquitos y Nauta para esclarecer las causas por las que un área con alta proporción de población en extrema pobreza ha tenido puntajes en las pruebas de suficiencia profesional menos deficientes que la casi totalidad de las regiones del país que disfrutaban de condiciones de menor deprivación socioeconómica.

#### *Promedios por grupo etario*

Para este análisis hemos dividido a la población en cuatro grupos de edad, como lo muestra la tabla n.º 17. El primero corresponde al intervalo de 21 a 25 años, el segundo al grupo comprendido entre los 26 y 30 años, el tercero al intervalo de 31 a 40 años y el cuarto al intervalo abierto de 41 años a más. Es visible que las pruebas de suficiencia profesional muestran una relación inversa entre la edad y el rendimiento académico en tanto que el intervalo más joven (21-25 años) tiene el promedio más alto y el intervalo de mayor edad (41 a más años) el promedio más bajo, esto es, 13,21 y 10,931, respectivamente. También es destacable que la pendiente es gradual, como corresponde a una relación inversa. La situación es muy semejante cuando dividimos a la población por género con la diferencia de que el grupo de menor rendimiento, por una diferencia de centésimos, pasa a ser el segmento de 31 a 40 años, en el caso de los varones.

Los resultados antes descritos merecen especial atención debido a que el reglamento general del concurso (D.S. 065-2001-ED) dio un peso desproporcionado a la edad al asignar 40 puntos a la experiencia docente. Ésta se midió en función de documentos que acreditaban la asistencia a seminarios, talleres y cursos breves; sin embargo se sabe que en estos espacios formativos está totalmente ausente la evaluación y para obtener las constancias o certificaciones es suficiente firmar el parte de asistencia. Como es obvio, un docente perteneciente al grupo etario más joven difícilmente puede competir, por razones temporales, con un docente mayor de 40 años en número de certificaciones y constancias que acrediten un cierto tipo de activismo magisterial. Este desacierto reglamentario, de tendencia gerontocrática, fue claramente contraproducente en

un país con población mayoritariamente joven, porque dio lugar a que se nombre en plazas estables de manera preferencial a los docentes con conocimientos más pobres y con menos posibilidades para remediar dicha deficiencia en el futuro.

En el momento en que redacto este trabajo, el Ministerio de Educación está tramitando un proyecto de ley de la carrera pública magisterial que además de no incluir concurso público alguno para el ingreso a la docencia, refuerza la tendencia gerontocrática porque los candidatos a nombramiento todo lo que tienen que presentar para su postulación es un conjunto de documentos cuya autenticidad, por añadidura, ha resultado frecuentemente cuestionable.

La diferencia entre los promedios del segmento más joven y del segmento más maduro es de 2,2790. Aunque esta cifra no es pronunciada, sin embargo es relevante en un concurso en el que se ganó una plaza por una diferencia de milésimos. Este valor ha sido bastante mayor en ocho sedes de administración de las pruebas. En la sede 03 de Breña la diferencia fue de 3,3580; en la sede 07 de San Borja, de 3,9931; en la sede 02 del Rímac, de 4,288 y en la sede 28 de la Dirección Regional de Cajamarca fue de 4,7949. En las siguientes cuatro sedes la diferencia fue mayor de 5 puntos. En la sede 11 de Cajatambo fue de 5,3047; en la sede 15 de Huarochirí, de 5,3154; en la sede 30 de la subregión Cutervo, de 5,1381 y en la sede 43 de San Lorenzo la diferencia fue de 8,5833, como puede verificarse al examinar la tabla n.º 18.

#### *Promedios según nivel y tipo de título*

En las tablas n.ºs 7 y 12 puede observarse que considerando a la población en su conjunto y dividiéndola según los niveles de nuestro sistema educativo (educación inicial, educación primaria y educación secundaria) han sido los docentes de educación inicial los que han obtenido los más altos promedios. En el caso de los graduados en institutos superiores pedagógicos la diferencia respecto de los docentes de educación primaria y de los docentes de educación secundaria fue de 3,9410 y 3,6900, respectivamente. Y en el caso de los graduados en las universidades como licenciados la diferencia respecto de los docentes de educación primaria y de educación secundaria ha sido de 5,8535 y 4,3308, respectivamen-

te. Es subrayable que en este último caso la diferencia en puntaje respecto de los docentes de educación primaria ha sido prácticamente equivalente al valor de una desviación estándar (ó) respecto del promedio general de la prueba A que aparece en la tabla n.º 7. De otra parte, los licenciados universitarios han alcanzado promedios mayores que los docentes titulados en los institutos superiores pedagógicos. Esta diferencia fue marcada en los casos de los docentes de educación inicial y de educación secundaria en los que los valores de las diferencias fueron 3,2295 y 2,5887, respectivamente.

Los resultados estadísticos anteriores sugieren que la formación académica proporcionada por las universidades tiende a ser mejor que la que brindan los institutos superiores pedagógicos. Esta tendencia se aprecia con nitidez en los niveles inicial y secundario.

Es el caso aclarar que el mejor promedio de los docentes de educación inicial no significa necesariamente que estén adecuadamente formados, pues su promedio general de 16,221 dista en 5,7790 puntos del puntaje mínimo aprobatorio de 22, equivalente a 27,5 aciertos. Ciertamente lo deseable hubiera sido que los docentes en promedio fuesen capaces de resolver el 70% de una prueba, lo que hubiera requerido que alcancen una nota de 28 puntos equivalente a 35 aciertos. Nos permitimos aclarar que, por razones que hemos expuesto en el párrafo referido a la administración de las pruebas, el valor del puntaje difiere del número de aciertos, los mismos que tienen un valor máximo de 50, concordante con el número de preguntas de cada una de las pruebas. Sin embargo, la diferencia a favor de los docentes de educación inicial justifica un estudio adicional para hacer inteligibles sus causas y poder deducir propuestas que podrían servir para mejorar, en general, la calidad de la formación docente.

### **Comportamiento revelador de algunas preguntas**

En lo que sigue procederé a analizar y comentar el comportamiento de algunas preguntas que formaron parte de las pruebas de matemática (D), comunicación (G) y de ciencias naturales (E). He elegido estas áreas por la centralidad que tienen no sólo en la formación integral de los docentes y de los escolares sino también

porque funcionan como ejes en los estudios que realizan comparaciones internacionales sobre la calidad de los sistemas educativos. Comenzaré entonces este análisis transcribiendo la pregunta n.º 30 de la prueba de matemática que se administró en la segunda fecha a 3.315 docentes.

- *Se dan las siguientes afirmaciones sobre polinomios*
  - 1)  $P(x)=x^2+3$  es irreducible sobre  $Q$
  - 2) Un polinomio es constante si es de grado cero o es el polinomio nulo
  - 3) Al sumar dos polinomios del mismo grado, se obtiene un polinomio que también es del mismo grado.
  - 4) El cociente de dos polinomios es también otro polinomio

*Señale la alternativa que muestre la secuencia correcta de verdad (V) o falsedad (F)*

  - A. FVFF
  - B. VFVF
  - C. VVFF
  - D. VFVV
  - E. FFVF

El sentido de esta pregunta fue explorar si los docentes de matemática eran capaces de identificar y aplicar las propiedades básicas o estructurales de los polinomios, materia ésta de su enseñanza en los cursos de álgebra que forman parte de los programas de estudios oficiales para educación secundaria. Sin embargo, es necesario hacer la salvedad de que con la metodología psicologista de elaboración de los currículos aplicada por el Ministerio de Educación durante los últimos 10 años, podría ocurrir que esta temática no esté adecuadamente definida pese a que es un componente indispensable para que nuestros escolares logren aprendizajes mínimamente aceptables en matemática. El índice de dificultad de esta pregunta fue de 0,9595 lo que significa que el 96% de nuestros docentes no es capaz de discriminar qué propiedades generales son propias de un polinomio y cuáles no. Resulta, por tanto, una declaración lírica pretender que docentes con deficiencias como ésta puedan impartir, satisfactoriamente, enseñanza del álgebra elemental.

En la prueba, que se administró a 4.354 docentes en la primera fecha, se incluyó como preguntas n.º 27 y n.º 29 las que se transcriben a continuación:

\*Si  $\int (x^2 - y - z) dx = \frac{1}{3}x^3 - yx - zx + k$

¿Cuál de las siguientes es verdadera?:

A.  $\int (x^2 - y + z) dx = \frac{1}{3}x^3 - yx + zx + k$

B.  $\int (x^2 - y - z) dx = \frac{1}{3}x^3 - yx - zx + k$

C.  $\int (x^2 - y - z) dx = \frac{1}{3}x^3 - yx - zx + k$

D.  $\int (x^2 - y + z) dx = \frac{1}{3}x^3 - yx - zx + k$

E.  $\int (x^2 - y - z) dx = \frac{1}{3}x^3 - yx - zx + k$

- Supongamos que:  $-x, y, z$  son enteros negativos ¿Cuál es el valor de verdad de cada una de las siguientes proposiciones?

I.  $|x| + |y| + |z| = x + y + z$

II.  $|x| + |y| + |z| = -x - y - z$

III.  $|x| + |y| + |z| = x - y - z$

IV.  $|x| + |y| + |z| = x + y - z$

A. FVVF

B. FFFV

C. FFVF

D. VFFF

E. VFVF

La primera de estas dos preguntas tuvo un índice de dificultad de 1,00 lo que significa que ninguno de los 4.354 profesores de matemática examinados en la primera fecha fue capaz de responderla correctamente; ello permite deducir que el análisis matemático básico es una materia ajena a la formación especializada de esta población que no ha podido realizar un ejercicio, casi mecánico, de manejo de integrales. Es claro que este resultado expresa una insuficiencia muy severa en la formación profesional que imparten tanto los institutos superiores pedagógicos como las facultades de Educación. En sistemas educativos de la comunidad internacional, como es el caso de Argentina, se incluye en los últimos semestres de la educación secundaria la enseñanza de cursos que versan sobre el tradicional cálculo integral y diferencial y la geometría analítica. Asumiendo la hipótesis de que el Ministerio de Educación dispusiera la introducción del cálculo integral y diferencial en los currículos escolares para mejorar la enseñanza de la matemática en nuestros centros educativos y hacerla competitiva y equiparable con la que se imparte en la comunidad internacional, la interrogante sería a qué docentes podríamos recurrir. Ciertamente, carecería de utilidad alguna la supuesta capacitación de los profesores en currículos, en gestión de la educación y en la monserga constructivista que han constituido los contenidos del Plancad y del Planged. El análisis de la pregunta siguiente completa este panorama sombrío en tanto que tuvo un índice de dificultad de 0,8536, lo que implica que más del 85% de la población examinada carece de una comprensión clara del concepto de valor absoluto.

En lo referente a prueba de comunicación (G), nos ha parecido pertinente ofrecer al lector una imagen del comportamiento de la pregunta n.º 23 de la prueba aplicada en la primera fecha y de la pregunta n.º 24 de la prueba aplicada en la segunda fecha, cuyos textos, transcribimos a continuación:

- *El significado y el significante son componentes del signo lingüístico. En esta perspectiva,*
  - I. *la homofonía se relaciona con el significante.*
  - II. *la polisemia se relaciona con el significante.*
  - III. *la sinonimia se relaciona con el significado.*
  - IV. *la antonimia se relaciona con el significante.*

*Elija la alternativa correcta.*

- A. I y III
- B. II y IV
- C. III y II
- D. I y IV
- E. II y I

- *Los fonemas son unidades distintivas porque*
  - A. *influyen en la distinción de vocales y consonantes.*
  - B. *ayudan a distinguir la lengua hablada de la escrita.*
  - C. *poseen un significado impreciso e incompleto.*
  - D. *permiten diferenciar un signo lingüístico de otro.*
  - E. *su buena articulación da distinción a los hablantes.*

El índice de dificultad de la primera pregunta fue de 0,6671 y el de la segunda pregunta fue 0,7137. El primer valor significa que casi el 67% de los docentes titulados como especialistas en comunicación para educación secundaria no maneja suficientemente el concepto básico de signo lingüístico y sus relaciones con las dimensiones semántica, sintáctica y fonológica. El segundo valor significa que la misma situación se reproduce en más de 71% de los docentes con el concepto de fonema que, por su carácter básico e instrumental, debía ser comprendido y manejado satisfactoriamente por el 100% de los profesores que tienen a su cargo la dirección del aprendizaje del lenguaje.

Una situación más preocupante aún es la que sugieren las preguntas números 4, 8, 10 y 32 de la prueba de ciencias naturales (E), aplicada en la primera fecha, cuyos textos transcribimos a continuación:

- *La teoría de la Gran Explosión (Big Bang) sobre el origen del universo es apoyada por:*
  - I. *La detección de la radiación cósmica de fondo.*
  - II. *El origen termonuclear de la energía de las estrellas.*
  - III. *La diferencia de la abundancia relativa de los elementos en el universo.*
  - IV. *La expansión del universo.*

Elija la alternativa correcta

- A. I, III
  - B. II, IV
  - C. II, III
  - D. II
  - E. I, IV
- *La contaminación del aire producida por el parque automotor nacional se debe principalmente a la emisión de:*
    - A. *Anhídrido carbónico, óxidos de cloro y partículas de carbón.*
    - B. *Monóxido de carbono, anhídrido carbónico y óxidos de azufre.*
    - C. *Óxidos de nitrógeno, monóxido de carbono y partículas de carbón.*
    - D. *Monóxido de carbono, freones y ozono.*
    - E. *Anhídrido carbónico, metano, ozono y partículas de carbón.*
  - *Observe, atentamente, el cuadro estadístico siguiente:*

Cuadro 1  
Perú: Número de docentes de Educación

	Total país (1+2)	Estatal (1)	No Estatal (2)
Total		250.080	90.850
Inicial		22.056	14.352
Primaria		132.500	40.452
Secundaria		95.524	36.046

Fuente: Ministerio de Educación, 2002.

De la cifra se deduce:

- A. *La Educación Secundaria es el 50% de la Educación Primaria.*
- B. *La Educación Primaria Estatal es más del 50% del total.*
- C. *La Educación Inicial es superior al 10% del total nacional.*
- D. *La Educación no Estatal es casi el 50% del total.*
- E. *La Educación Secundaria privada es el rubro más pequeño de la Educación no Estatal.*

- Dadas las siguientes afirmaciones:
  - I. La energía cinética es la que adquiere un objeto cuando está en movimiento.
  - II. La energía potencial es la que tiene un objeto de acuerdo con su posición, su estado o su forma.
  - III. La energía interna depende tanto de la velocidad como de la altura del objeto.

Señale la alternativa que contiene sólo afirmaciones verdaderas.

- A. I, II y III
- B. I y II
- C. I
- D. I y III
- E. II y III

El índice de dificultad de la primera pregunta, para los 3.901 profesores de ciencias naturales que fueron examinados en la primera fecha, fue de 0,9628, lo que significa que menos de un 4% de la población ha comprendido mínimamente el significado conceptual de la teoría cosmológica de la gran explosión conocida como Big Bang. La segunda pregunta tuvo un índice de dificultad de 0,957, lo que implica que menos de un 5% de la población fue capaz de identificar correctamente los principales elementos contaminantes producidos por el parque automotor. La tercera pregunta tuvo un índice de dificultad de 0,9430, lo que revela que menos de un 6% de la población fue capaz de realizar una lectura comprensiva de una tabla de doble entrada que muestra porcentajes. La cuarta pregunta tuvo un índice de dificultad de 0,9274, lo que revela que casi el 93% de la población no comprende suficientemente el concepto físico fundamental de energía.

Las insuficiencias anteriores adquieren mayor severidad cuando se toma en cuenta que son los docentes de esta especialidad los que tienen a su cargo la enseñanza de la cosmología y del cuidado y conservación del ecosistema. Esperar que docentes con las insuficiencias cognoscitivas puestas de manifiesto por estas preguntas puedan ofrecer un servicio adecuado en la enseñanza de las materias antes mencionadas carece de fundamento razonable. Asimismo-

mo, el hecho de que pese a su supuesta formación profesional en ciencias básicas, los docentes no sean capaces de leer comprensivamente una tabla de doble entrada que muestra porcentajes gruesos, permite deducir que carecen de los instrumentos mínimos requeridos por el auto aprendizaje. Como se comprenderá, estos resultados puntuales se condicen simétricamente con los resultados estadísticos que presentamos en este informe y con los aportados por las pruebas UNESCO y PISA, antes mencionadas, todo lo cual cuestiona seriamente la calidad de los servicios que ofrecen los institutos superiores pedagógicos y las facultades de Educación.

### **Estadísticas de las pruebas**

Las afirmaciones estadísticas que se hacen en este informe pretenden ser válidas para la totalidad del magisterio nacional sobre la base del siguiente argumento: Aunque se ha utilizado una fracción no-aleatoria de la totalidad de los docentes del Perú, sus dimensiones, 95.219 examinados, constituyeron en el año 2002 más del 25% de los miembros del magisterio en servicio distribuidos en 60 ciudades del territorio nacional. Estas peculiaridades de la población examinada, su gran tamaño y su distribución en la totalidad del territorio nacional, nos permiten asegurar que las propiedades registradas por las pruebas de suficiencia profesional-2002, configuran la imagen más aproximada del grado de la calidad de la formación docente en el Perú lograda por investigación alguna hasta la fecha.

Ciertamente, esta pretensión no es incompatible con nuestro reconocimiento de que éste es sólo el primer esfuerzo, con importantes limitaciones, que demanda continuidad, correcciones, aplicaciones y profundización para alcanzar estándares aceptables de conocimiento en el sistema educativo nacional.

#### *Pruebas de educación inicial (A)*

Las pruebas de educación inicial (A) se administraron en todo el territorio nacional a 10.387 docentes, de los cuales 7.422 se evaluaron en la primera fecha (ver tabla n.º 1), en todas las sedes excepto

en la Use 17 correspondiente a la Dirección de Educación de Lima Metropolitana en la que no se registraron concursantes en esta prueba ni en ninguna de las otras con excepción de la prueba de educación artística (H). Los 2.965 docentes restantes correspondieron a la segunda fecha distribuidos en sólo 35 sedes por no haberse registrado concursantes de este género en las 25 sedes que se detallan en la tabla n.º 3. Las damas alcanzaron un promedio general, en el ámbito nacional, de 16,293 y los varones un promedio general de 11,4 como puede observarse en la tabla n.º 8. El promedio general para ambos géneros ha sido de 16,221 como puede apreciarse en la tabla n.º 7 y es más alto que el que han obtenido los concursantes en cada una de las trece pruebas restantes. La desviación estándar ó es de 5,958. Revela que el grupo era homogéneo; sin embargo, los varones con un promedio de 11,4 tienden a ubicarse alrededor de 0,8208 ó por debajo del promedio del grupo. Tanto en la primera fecha de administración como en la segunda, la prueba (A) presentó valores muy aceptables para su índice de discriminación y su índice de dificultad. En relación con el primero, como se muestra en la tabla n.º 20, los valores fueron 0,3143564 y 0,23949089, para la primera y segunda fecha, respectivamente, mientras que para el segundo los valores fueron de 0,4758906 y 0,55235076, para la primera y segunda fecha, respectivamente. Estas cifras, según la bibliografía especializada, muestran que la prueba tuvo capacidad adecuada para discriminar entre grupo de alto rendimiento y el de bajo rendimiento y que sus preguntas tuvieron un muy moderado índice de dificultad. El promedio más alto en las pruebas de tipo A lo obtuvo la sede 07 (San Borja, Lima) con un puntaje de 21,78. El promedio más bajo se registró en la sede 46 (Requena) con un puntaje de 9,73 (ver la tabla n.º 19).

### **Pruebas educación primaria (B)**

Las pruebas de educación primaria se administraron en todo el territorio nacional a 34.631 docentes, a 20.242 en la primera fecha y a 14.389 en la segunda fecha (ver tabla n.º 1). La población estuvo constituida por 24.752 damas y 9.733 varones.<sup>1</sup> El promedio gene-

<sup>1</sup> De aquí en adelante, cuando el autor se refiera al total de docentes que rindió algún tipo de prueba, estará incluyendo también a los que omitieron marcar su sexo y que figuran en la tabla n.º 2 en columna aparte. (N. del E.)

ral para la población es de 11,108 con una desviación estándar ó de 5,314. El promedio general de damas fue 11,37 y el promedio general de varones 11,819. La desviación estándar ó para las damas fue 5,275 y para los varones 5,511. Debemos destacar que los promedios de las damas y los varones obtenidos en las pruebas A son muy semejantes y que el grado de homogeneidad de los géneros según lo mostrado por la tabla n.º 8 es prácticamente el mismo.

En lo referente al índice de dificultad en la primera fecha su valor fue de 0,78839838 y en la segunda 0,57978456, lo que significa que las pruebas administradas inicialmente fueron difíciles mientras que las posteriores tuvieron un grado de dificultad muy moderado. El índice de discriminación de la primera fecha, 0,04805269, fue muy inadecuado y afecta la validez de la pruebas, sin embargo, en la segunda fecha el índice de discriminación de 0,29202448 es de un valor bastante aceptable y compensa en alguna medida la deficiencia de la primera fecha, la misma que es muy explicable si se considera el tiempo extremadamente breve dedicado a la planificación y elaboración de las pruebas (menos de siete semanas). El máximo promedio en el ámbito nacional en las pruebas del tipo B lo obtuvo la sede 11 (Cajatambo) con un puntaje de 17,08; el promedio mínimo se registró en la sede 46 (Requena) con un puntaje de 7,36. Sin embargo, es importante señalar que en la sede 11 de Cajatambo solamente 18 personas se sometieron a las pruebas del tipo B, lo que debilita la significación estadística de su promedio.

#### *Pruebas de educación especial (C)*

Las pruebas de educación especial se administraron en todo el territorio nacional a 576 docentes, de los cuales 488 pertenecían al género femenino y 80 al masculino (ver tabla n.º 2). El promedio general de la población fue de 9,745 con una desviación estándar ó de 5. El promedio de las damas ha sido 9,967 y el de los varones 8,79. En la primera fecha el índice de discriminación de las pruebas del tipo C fue de 0,0262007 y el índice de dificultad fue de 0,84314928. Como en el caso anterior, en la primera fecha el instrumento careció de poder discriminativo significativo y tuvo un alto grado de dificultad. Dichos valores se modificaron positivamente en la segunda fecha, pues en esta oportunidad la prueba C tuvo un

índice de discriminación significativo de 0,30173913 y un índice de dificultad moderado, 0,61816092. Asimismo, el cálculo del promedio de rendimiento, en esta prueba, en algunas sedes resultó una ficción estadísticamente inadmisibles, pues en ocho sedes dio examen sólo un docente en cada una de ellas y en 19 sedes dieron examen cuatro personas o menos en cada una de ellas. Esto significa que la estrategia de introducir un factor de corrección en función del rendimiento promedio de una determinada sede ha dado lugar a que se asigne puntaje aprobatorio a numerosas personas que obtuvieron realmente cuatro puntos o menos sobre un máximo de 50. Ciertamente existe el riesgo de que estas personas estén disfrutando actualmente de la estabilidad de un nombramiento. Por otro lado, la sede 32 (Cusco) obtuvo el máximo promedio en el ámbito nacional, 12,78, en las pruebas de tipo C. El promedio mínimo se registró en la sede 59 (Tumbes) con un valor numérico de 3,96.

*Pruebas educación secundaria, matemática (D)*

Las pruebas del tipo (D) se administraron a 7.679 docentes, de los cuales 3.329 fueron damas y 4.303 varones. El rendimiento promedio de esta población fue de 10,228 y la desviación estándar ó tuvo el valor de 4,512. El promedio de las damas fue 10,038 con una desviación estándar ó de 4,353. El promedio de los varones es de 13,009 con una desviación estándar ó de 5,307. Ésta es el área de especialización en la que los varones obtuvieron una diferencia de casi tres puntos sobre el promedio de rendimiento de las mujeres.

En la primera fecha el índice de discriminación de las pruebas tipo A fue de 0,22442468 y el índice de dificultad de 0,6465. En la segunda fecha el índice de discriminación incrementó su valor obteniendo a 0,24637939 y el índice de dificultad decreció a 0,61936048. De acuerdo con la bibliografía especializada, las pruebas tipo C han tenido características muy aceptables en el sentido de que los valores de sus índices de discriminación se aproximan notablemente al valor deseable de 0,25 y su índice de dificultad, en ambas administraciones, fue moderado. Asimismo, la sede que obtuvo el máximo promedio en el ámbito nacional en las pruebas de tipo C fue la n.º 41 (Iquitos) con un promedio de 13,13 y la que obtuvo el promedio mínimo fue la sede n.º 46 (Requena) con un puntaje de 6,69.

#### *Pruebas de educación secundaria, comunicación (E)*

Las pruebas de tipo E se administraron a 6.780 docentes, de los cuales 3.959 eran damas y 2.805 varones (ver la tabla n.º 2). El promedio general de esta área fue de 12,1727 con una desviación estándar ó de 5,188. Las damas obtuvieron el promedio de 12,873 con una desviación estándar ó de 5,101 y los varones registraron un promedio de 10,4 con una desviación estándar ó de 4,619. En la primera fecha las pruebas de comunicación tuvieron un índice de dificultad de 0,921286685, cifra que las tipifica como especialmente difíciles. Esto dio lugar a que su índice de discriminación haya sido negativo, -0,00089667, lo que ciertamente afecta su validez de constructos. Esta deficiencia fue corregida en la segunda fecha debido a la especial atención que el equipo de trabajo dedicó a esta prueba a causa de que se contó con informes del personal encargado de la administración de las pruebas que justificaron esta decisión. En la segunda fecha la mejora fue sorprendente porque tanto el índice de discriminación como el índice de dificultad alcanzaron valores que de acuerdo con la bibliografía especializada son muy adecuados, 0,48277147 y 0,54197497, respectivamente. Esto significa que en la segunda administración las pruebas del tipo E tuvieron un grado de dificultad muy moderado y separaron adecuadamente a los concursantes de alto rendimiento de los de bajo rendimiento. En estas pruebas el máximo promedio en el ámbito nacional lo obtuvo la sede n.º 41 (Iquitos) con un puntaje de 16,99 y el promedio mínimo lo obtuvo la sede 57 (Tocache) con un puntaje de 8,36.

#### *Pruebas de educación secundaria, ciencias sociales (F)*

Las pruebas del tipo F se aplicaron a 9.205 docentes, 4.904 damas y 4.267 varones, según muestra la tabla n.º 2. El promedio general en esta área de especialización fue 11,56 con una desviación estándar ó de 4,87. El promedio de las damas fue de 11,268 con una desviación estándar ó de 4,808 y el promedio de los varones de 11,905 con una desviación estándar ó de 4,15. En la primera administración el índice de discriminación fue de 0,247421109 y el índice de dificultad de 0,59987292. Los valores anteriores significan

que este tipo de prueba desde su primera administración mostró una aceptable capacidad discriminativa para diferenciar a los concursantes de alto rendimiento de aquellos de bajo rendimiento. El grado de dificultad de las preguntas fue muy moderado y estuvo por tanto al alcance de los concursantes. En la segunda administración el índice de discriminación mejoró ligeramente al alcanzar el valor de 0,26834104. El índice de dificultad se mantuvo prácticamente constante con un valor de 0,58997361. En las pruebas F la sede 07 (San Borja-Lima) alcanzó en el ámbito nacional el promedio más alto con un puntaje de 14,17. El promedio más bajo lo obtuvo la sede 27 (Sara Sara-Cora Cora) con un puntaje de 7,93.

*Pruebas de educación secundaria, ciencias naturales (G)*

Las pruebas del tipo G se aplicaron a 9.498 docentes, 5.692 damas y 3.766 varones como lo muestra la tabla n.º 2. El promedio de esta área de especialización fue 9,854 con una desviación estándar ó de 5,037. Las damas obtuvieron un promedio de 9,934 con una desviación estándar ó de 4,98 y los varones registraron un promedio de 9,739 con una desviación estándar ó de 5,123. En la primera administración las pruebas G obtuvieron un índice de discriminación de 0,27309855 y un índice dificultad de 0,65742698, lo que significa que los instrumentos de este tipo estuvieron dotados de capacidad discriminativa adecuada y de un grado de dificultad moderado. En la segunda administración el índice de discriminación de 0,25450436 y el índice de dificultad de 0,618993133 son valores que expresan que la tendencia observada en la primera fecha se mantuvo. En breve, las pruebas del tipo G como las del tipo F, las del tipo D y las del tipo A tuvieron un comportamiento estadístico que abona a favor de su validez y confiabilidad. La sede 41 (Iquitos) obtuvo en estas pruebas el máximo promedio en el ámbito nacional con un puntaje de 15,42. El promedio mínimo lo registró nuevamente la sede 27 (Sara Sara-Cora Cora) con un puntaje de 5,98.

*Pruebas de educación secundaria, educación artística (H)*

Las pruebas del tipo H se aplicaron a 2.188 docentes, 918 damas y 1252 varones como se registra en la tabla n.º 2. El promedio general

de esta población fue 9,295 y fue el más bajo en relación con los trece restantes. La desviación estándar (s) fue de 4,155. Las damas obtuvieron un promedio de 9,171 con una desviación estándar (s) de 4,1267. Los varones obtuvieron un promedio de 9,382 con una desviación estándar (s) de 4,081. En la primera administración las pruebas de tipo H tuvieron un índice de discriminación de 0,22031426 y un índice de dificultad de 0,64617415. En la segunda administración el índice de discriminación decreció a 0,217573 y el de dificultad alcanzó 0,66405694. Los valores anteriores expresan que las pruebas del tipo H estuvieron en lo referente al índice de discriminación muy próximas a los valores aceptables y en relación con el índice de dificultad tuvieron un valor que las ubica dentro de un grado moderado de dificultad. El promedio más alto en las pruebas de tipo H lo obtuvo la sede 6 (Ate-Vitarte) con un puntaje de 12,89. El promedio mínimo lo registró la sede 18 (Amazonas) con un puntaje de 6,56.

Es importante señalar que en el caso de las pruebas de tipo H también se han producido situaciones en las que ha carecido de significado estadístico calcular el promedio por sede debido a la cantidad ínfima de concursantes en 12 de ellas. Así, aportando algunos detalles, señalamos que en las sedes 11 (Cajatambo), 14 (Oyón), 17 (Dirección Regional de Lima), 23 (Abancay), 27 (Sara Sara-Cora Cora), 29 (Chota), 30 (Cutervo), 42 (Alto Amazonas), 44 (Nauta), 47 (Contamana) y 57 (Tocache) se sometieron a la prueba H tres personas o menos en cada una de ellas.

#### *Pruebas de educación secundaria, de idioma inglés (I)*

En las tablas estadísticas que aparecen en este estudio, las pruebas I figuran bajo el rubro de idioma extranjero. Sin embargo, como el único idioma extranjero que se enseña en nuestros centros educativos públicos de nivel básico es el inglés y como las pruebas del tipo I versaron solamente sobre esta lengua, hemos preferido llamarlas directamente por su nombre en el subtítulo que encabeza este párrafo. Las pruebas de tipo I fueron administradas en todo el territorio nacional a 809 docentes, 631 damas y 177 varones como se aprecia en la tabla n.º 2. El promedio de esta área de especialización fue de 13,899 con una desviación estándar (s) de 6,296. Las

damas obtuvieron un promedio de 14,157 con una desviación estándar (s) de 6,122 y los varones alcanzaron un promedio de 12,976 con una desviación estándar (s) de 6,834. En la primera fecha las pruebas del tipo I tuvieron un índice de discriminación de 0,35759322 y un índice de dificultad de 0,56283186. Estos valores son muy aceptables de acuerdo con la bibliografía especializada y revelan que las pruebas de tipo I integraron el conjunto de los instrumentos que cumplieron con el objetivo de diferenciar a los concursantes de alto rendimiento de los de bajo y estuvieron dentro de la capacidad de respuesta de los concursantes con un índice muy moderado de dificultad.

En la segunda administración se consolidó la tendencia anterior en tanto que el índice de discriminación alcanzó 0,38473913 y el índice de dificultad de 0,53008403. El promedio más alto en el ámbito nacional en las pruebas de tipo I lo obtuvo la sede 40 (Callao) con un puntaje de 21,29. El promedio más bajo se registró en la sede 36 (Paracas-Chincha) con un puntaje de 5,69. En este caso también se repitieron situaciones que evidenciaron que el cálculo del promedio por prueba y por sede no tuvo significación estadística. Aportando detalles, en las sedes n.º 8 (Cañete), 10 (Huaral), 14 (Oyón), 18 (Chachapoyas), 22 (Pomabamba), 24 (Andahuaylas), 26 (Huamanga), 42 (Alto Amazonas), 48 (Madre de Dios), 51 (Piura), 55 (Tarapoto), 56 (Mariscal Cáceres) y 60 (Pucallpa) se sometieron a examen en cada una de ellas dos personas o menos con la excepción de Andahuaylas, sede en la que dieron examen tres docentes.

#### *Pruebas de educación física (J)*

Las pruebas de tipo J se aplicaron en todo el territorio nacional a 4.921 docentes, 1126 damas y 3758 varones como puede verse en la tabla n.º 2. La población de esta área de especialización obtuvo un promedio general de 11,282 con una desviación estándar (s) de 5,146. Las damas obtuvieron un promedio de 10,877 con una desviación estándar (s) de 5,16. Los varones registraron un promedio de 11,434 con una desviación estándar de (s) 5,133. En la primera fecha las pruebas del tipo J obtuvieron un índice de discriminación de 0,26720217 y un índice de dificultad de 0,61700402. En la segunda administración el índice de discriminación fue de

0,26376104 y el índice de dificultad de 0,57223131. Los valores anteriores permiten interpretar que las pruebas del tipo J, en ambas fechas, tuvieron un índice de discriminación aceptable que se mantuvo prácticamente constante y un índice de dificultad que osciló entre moderado y muy moderado. La estabilidad de estos valores abona en favor de la validez y la confiabilidad de las pruebas. La sede 41 (Iquitos) obtuvo el promedio más alto en el ámbito nacional en las pruebas J con 15,28 de puntaje. El promedio mínimo lo obtuvo la sede 27 (Sara Sara-Cora Cora) con un puntaje de 7,51.

*Pruebas de educación secundaria, variante técnica (K)*

Las pruebas del tipo K se administraron en todo el territorio nacional a una población de 4.967 docentes, 2.190 damas y 2.756 varones, como muestra la tabla n.º 2. Esta población alcanzó un promedio general de 12,502 con una desviación estándar (s) de 5,137. Las damas obtuvieron un promedio de 12,497 con una desviación estándar (s) de 4,34 y los varones consiguieron un promedio de 12,519 con una desviación estándar (s) 5,292. En la primera administración, las pruebas del tipo K obtuvieron un índice de discriminación de 0,27578403 y un índice de dificultad de 0,58563286. En la segunda administración se registró como índice de discriminación 0,25468944 y como índice de dificultad 0,57216833. Como puede observarse en la tabla n.º 20 los valores antes mencionados muestran una estabilidad significativa que pone en evidencia que estas pruebas tuvieron capacidad discriminativa muy aceptable y un grado de dificultad muy moderado, lo que abona a favor de su validez y confiabilidad. La sede 16 (Barranca) es la que obtuvo en estas pruebas el promedio más alto en el ámbito nacional con un puntaje de 17,28. El promedio más bajo lo obtuvo la sede 26 (Ayacucho) con un puntaje de 7,90.

*Pruebas para centros de educación ocupacional (L)*

Las pruebas del tipo L se administraron en el territorio nacional a 1.093 docentes, 730 damas y 358 varones como muestra la tabla n.º 2. El promedio general de esta población fue de 10,247 con una desviación estándar (s) de 4,799. Las damas obtuvieron un promedio

de 9,826 con una desviación estándar (s) de 4,722 y los varones alcanzaron un promedio de 11,109 con una desviación estándar (s) de 4,852. En la primera administración las pruebas de tipo L tuvieron un índice de discriminación de 0,25121639 y un índice de dificultad de 0,63940814. En la segunda administración el índice de discriminación decreció a 0,23173333 y el índice de dificultad también disminuyó a 0,58836879. Los valores anteriores indican que tanto las pruebas de la primera administración como las de la segunda se encuentran dentro de parámetros aceptables, sin embargo, puede afirmarse que el comportamiento de las pruebas de la primera fecha fue teóricamente mejor que el de las pruebas de la segunda administración debido a que tuvieron un mayor poder discriminativo pese a poseer un mayor grado de dificultad dentro de la escala que hemos asumido en este trabajo. La sede 41 (Iquitos) logró en estas pruebas el más alto promedio en el ámbito nacional con un puntaje de 14,60. El menor promedio lo obtuvo la sede 35 (Ica) con un puntaje de 4,72. En el caso de estas pruebas también hay que señalar que ha habido 11 sedes, sede 11 (Cajatambo), 14 (Oyón), 21 (Chimbote), 22 (Pomabamba), 29 (Chota), 31 (Jaén), 36 (Paracas-Chincha), 48 (Madre de Dios), 55 (Tarapoto), 56 (Mariscal Cáceres) y 57 (Tocache) en las que el cálculo de los promedios por sede ha carecido de significado estadístico debido a que en cada una de ellas se sometieron a las pruebas tres personas o menos.

#### *Pruebas para institutos tecnológicos (M)*

En primer lugar cabe aclarar que las pruebas del tipo M fueron las únicas que se aplicaron a docentes ubicados en el nivel de la educación superior, los mismos que totalizaron 1.985, 572 damas y 1409 varones. Fue un acierto importante que la dirección correspondiente del Ministerio de Educación no se opusiera a la evaluación, como sí lo hizo la Dirección Nacional de Formación y Capacitación Docente (Dinfocad) a cargo de la conducción de los institutos superiores pedagógicos. Es oportuno resaltar que estos docentes actúan en un nivel estrechamente ligado al desarrollo tecnológico del país y, por ende, a su desarrollo social y económico, con una población de 262.832 estudiantes matriculados según cifras oficiales actualizadas al año 2003. El promedio general obte-

nido por esta población fue de 12,753 con una desviación estándar de 4,65. Las damas obtuvieron un promedio de 12,531 con una desviación estándar (s) de 4,42 y los varones un promedio de 12,835 con una desviación estándar (s) de 4,741. En la primera administración las pruebas de tipo M obtuvieron un índice de discriminación de 0,26517639 y un índice de dificultad de 0,56849273. En la segunda administración los valores obtenidos para los índices de discriminación y dificultad fueron de 0,23970485 y de 0,59412112 respectivamente. Los valores antes mencionados muestran que la capacidad discriminativa de la prueba fue ligeramente mejor en la primera fecha que en la segunda y que el índice de dificultad en ambos casos estuvo dentro de los parámetros que lo tipifican como muy moderado. La sede 25 (Arequipa) alcanzó el mayor promedio en el ámbito nacional con un puntaje de 14,20. El menor puntaje lo obtuvo la sede 35 (Ica) con un puntaje de 6,47.

#### *Las pruebas de educación religiosa (N)*

Las pruebas de tipo N, preparadas, como hemos señalado antes, directamente por la Comisión Episcopal, se aplicaron en todo el territorio nacional a 488 docentes, 325 damas y 162 varones. El promedio general obtenido por esta población fue de 12,94 con una desviación estándar (s) de 6,537 que es la de mayor valor como lo muestra la tabla n.º 7. Esta última cifra significa que el rendimiento de los docentes de educación religiosa ha sido el más heterogéneo, considerando las 14 áreas exploradas. Las damas obtuvieron un promedio de 12,416 con una desviación estándar (s) de 6,827 y los varones consiguieron un promedio de 14,02 con una desviación estándar (s) 5,81. Las pruebas del tipo N en la primera administración tuvieron un índice de discriminación de 0,24212836 y un índice de dificultad de 0,5555642. En la segunda administración el índice de discriminación se incremento notablemente a 0,44517241 y el índice de dificultad a 0,56813853. Estos valores abonan significativamente a favor de la validez y confiabilidad de las pruebas en tanto que muestran una capacidad creciente para diferenciar el grupo de alto rendimiento del grupo de bajo rendimiento y una estabilidad en el grado de dificultad de las preguntas. La sede 01 (San Juan de Miraflores) obtuvo el mayor

promedio en el ámbito nacional con un puntaje de 19,25 y la sede 35 (Ica) registró el promedio más bajo de las pruebas N y de los 14 tipo de pruebas con un puntaje de 0,20. Este último puntaje es el promedio de 20 docentes que fueron examinados en la ciudad de Ica.<sup>2</sup>

<sup>2</sup> La sede 38 en la prueba N fue afectada presumiblemente por un error en el reparto de los cuadernillos de la prueba que dio lugar a que la calificación se realice con una clave incorrecta. El error fue rectificado posteriormente y el promedio de esta sede cuando fue recalificada fue 18,87, puntaje presumiblemente sesgado debido a que probablemente algunos concursantes conocieron el texto del cuadernillo que ya había sido distribuido en la primera fecha y que se les entregó, en la segunda fecha, por error. Hemos mantenido el valor 0,20 porque el cambio no resolvería la deficiencia antes descrita.

## Conclusiones

1. La experiencia de aplicación de las pruebas de suficiencia profesional-2002 para medir los conocimientos especializados, pedagógicos y la denominada cultura general de los concursantes al nombramiento en plazas docentes en centros educativos estatales ha constituido una acción evaluativa sin precedentes en el Perú y que, hasta la fecha de la redacción de este trabajo, tiene todas las características de un esfuerzo aislado que no ha ocupado lugar alguno en la políticas del Ministerio de Educación después de junio del año 2002. Por el contrario, inmediatamente después de ser reemplazado el ministro Nicolás Lynch, su sucesor procedió a desactivar la oficina que planificaba una nueva evaluación para los docentes que debieron participar en un concurso para cubrir plazas estables en áreas rurales.

2. Desde el punto de vista de los lineamientos de política educativa que han orientado los programas de capacitación docente ejecutados por el Ministerio de Educación del Perú, la situación académica de los profesores mostrada por los indicadores estadísticos contenidos en este informe puede interpretarse como un fracaso social y financiero del Programa Nacional de Capacitación Docente (Plancad) realizado durante la última década. Este resultado es explicable si se considera que el referido programa enfatizó la capacitación de 180.787 docentes en áreas tales como planeamiento estratégico, gestión institucional y corrientes pedagógicas, supuestamente innovativas, como el constructivismo de Piaget, Vygotski y Brunner, en versiones de divulgación de segunda y tercera mano. La exacerbación de esta «nueva propuesta pe-

dagógica» dio lugar a que en los hechos no se concediera, prácticamente, espacio alguno a la capacitación de los docentes en disciplinas especializadas como matemática, lenguaje, física, química, biología, psicología, lógica, filosofía, economía, etc., las mismas que constituyen el *qué* de la educación, dimensión que es condición necesaria para que el docente esté en aptitud de impartir buena enseñanza y dirigir adecuadamente el aprendizaje. En cuanto al financiamiento, el actual Ministro de Educación en su informe al Congreso de la República, en septiembre del año 2004, ha reconocido como deuda sólo con el Banco Interamericano y el Banco Mundial, incluyendo compromisos hasta el año 2007, la suma de 625.567 millones de dólares americanos. No se conoce públicamente los montos gastados con cargo al tesoro público ni con cargo a los préstamos provenientes de la fundación alemana KfW. Sin embargo, cálculos no oficiales permiten conjeturar con fundamento que el monto total de la deuda externa contraída por la ejecución, principalmente del Plancad y del Planged, es aproximadamente 1.000 millones de dólares americanos, lo que significaría que el costo unitario de estos programas fallidos bordearía los 5.000 dólares americanos por docente, gastados, en proporción muy alta, a través de la modalidad de terciarización de los servicios. Corroboramos nuestra evaluación negativa de dichos programas el hecho de que los escolares peruanos hayan ocupado el último lugar de la región en las pruebas de la UNESCO cuyos resultados se publicaron el año 2001 y el último lugar en las pruebas PISA administradas a los escolares de 15-16 años de 41 países del planeta.

3. El perfil general de la formación docente en el Perú mostrado por los muy bajos rendimientos académicos de 95.219 docentes peruanos en las pruebas de suficiencia profesional-2002, segmentadas en 14 áreas de especialización, permite deducir que no sólo los programas de capacitación docente están severamente desorientados sino que, principalmente, el sistema de formación docente constituido por los institutos superiores pedagógicos, dirigidos académicamente por Dinfocad, y las facultades de Educación de las universidades atraviesa por una crisis cuyos inicios tienen no menos de 25 años de antigüedad. Los síntomas más ostensibles de esta crisis son el «pedagogicismo» que afecta a los centros de formación magisterial y, su complemento, el gremialismo corporativista que se expresa en la trayectoria del sindicato nacio-

nal de docentes. El pedagogicismo consiste en la tendencia a construir los planes de estudios de los centros de formación docente concediendo un 70% de las horas de clase a los cursos pedagógicos que constituyen *el cómo* de la educación y solamente un 30% a los cursos sobre disciplinas científicas especializadas, las cuales constituyen *el qué* de la educación. Esta estructura del plan de estudios de formación docente es obsoleta y no se utiliza hace más de tres décadas en los países desarrollados debido a que existe plena conciencia de que con la complejización y enriquecimiento de la ciencia y de la tecnología, la precondition necesaria para ejercer la docencia eficientemente es conocer primero el contenido y organización lógica de lo que se pretende enseñar para luego aprender, dando preferencia a la práctica en el aula, los procedimientos más adecuados para enseñar aquello que previamente se conoce. La paradoja praxiológica que se refleja duramente en los muy precarios aprendizajes de los escolares peruanos consiste en que sus profesores son formados durante cinco años para que intenten «enseñar» lo que, estrictamente, no saben.

El gremialismo corporativista es el correlato en el mercado laboral y en la legalidad imperante del pedagogicismo. Se manifiesta en el hecho de que los contratos y nombramientos para prestar servicios en los centros educativos dependientes del Ministerio de Educación excluyen drásticamente la incorporación de psicólogos, lingüistas, matemáticos, traductores de lengua extranjera, físicos, ingenieros, etc., a la enseñanza, prohibición que priva a la educación pública de los profesionales que mejor conocen los contenidos de la enseñanza, las teorías y metodologías científicas que permiten explicar y regular el aprendizaje humano. Esta exclusión, de la que están libres los centros educativos privados, se sustenta en los textos de la ley del profesorado, la ley de educación vigente n.º 24088 y la ley de colegiatura de los miembros del magisterio nacional, normas legales que, irónicamente, son invocadas por las dirigencias sindicales como conquistas en beneficio de la educación nacional.

4. La no existencia en el Perú de una cultura de respeto a los conocimientos científicos y a la evaluación se manifestó claramente en la resistencia que opusieron a la aplicación de las pruebas de suficiencia profesional las autoridades tradicionales del Ministerio de Educación, que mantuvo intocadas el ministro Nicolás Lynch,

posiblemente debido a lo efímero de su ejercicio, el sindicato docente denominado SUTEP y la clase política. La costumbre inveterada de hacer uso del poder para distribuir los cargos públicos en función de intereses divorciados del interés del país, permite entender que a ninguno de los sectores que tiene voz y voto en el reparto de los cargos públicos le interese un mecanismo de asignación de plazas docentes en función estricta y excluyente de la capacidad profesional de los postulantes, evaluada mediante procedimientos científicos impersonales.

5. La inexistencia de una cultura de la evaluación de desempeños y no de documentos administrativos, la misma que es el sustento de una sociedad meritocrática, se manifestó en el hecho de que el Perú es uno de los escasos países de la región que carece de un sistema de pruebas nacionales periódicas para medir los rendimientos académicos de los estudiantes de educación primaria, secundaria y superior. Estos sistemas existen hace más de 40 años en Chile, Brasil, Colombia, Argentina y en otros países de Sudamérica, lo que ha dado lugar a que la población de nuestros vecinos esté habituada desde la infancia a los procesos de evaluación académica y a que las decisiones sobre políticas educativas, asignación de recursos y estímulos se tomen teniendo en cuenta los resultados de las evaluaciones nacionales, pues éstas constituyen no sólo preocupación del sector Educación sino, principalmente, de las familias y de la población en su conjunto. La constatación de este hecho basta para explicar por qué el Perú ocupa el último lugar en rendimiento académico entre los países de la región de América Latina y el Caribe y entre los países del planeta, según las pruebas internacionales UNESCO (1998-2000) y las pruebas PISA. También es suficiente para comprender por qué cualquier proyecto nacional meritocrático, en las condiciones actuales, es simplemente una quimera que colisiona con la contracultura imperante que privilegia, como mecanismo para lograr un cargo público, el manejo de las relaciones familiares, amicales, políticas, de filiación religiosa, sindicales, etc., y/o la negociación efectiva de dinero y servicios inficionados por la corrupción, declarativamente descalificada pero, en los hechos, muy bien conservada por la institucionalidad.

6. El promedio general de 11,9230 permite afirmar que el rendimiento académico global de la población es ligeramente mayor que la mitad del puntaje aprobatorio esperable, estimable en aproximadamente 20 puntos considerando un puntaje máximo de 40. Si se asume la hipótesis de que las pruebas pretendían medir los conocimientos mínimos *que deberían* poseer los docentes para cumplir satisfactoriamente sus funciones, entonces el promedio general se ubica aproximadamente a una distancia de  $3/4$  del puntaje ideal teórico. En los valores obtenidos para la confiabilidad mediante la fórmula conocida como Kuder Richardson, 20 es aceptable. La plausibilidad de estos cálculos está corroborada por los valores del error estándar que en todos los casos, por ser menor que 4, contribuye a corroborar la hipótesis que afirma que los instrumentos utilizados son aceptablemente estables. La validez predictiva de las pruebas es uno de los asuntos abiertos a la investigación y, seguramente, una de las prioridades en la situación contractual de que el Ministerio de Educación decida crear un sistema de pruebas nacionales para supervisar la calidad de la formación profesional que brindan los institutos superiores pedagógicos y las universidades; ello para garantizar el acceso a las plazas docentes mediante mecanismos transparentes, científicos y meritocráticos. Los índices de discriminación y de dificultad de las pruebas permiten conjeturar su capacidad para diferenciar al grupo de alto rendimiento del grupo de bajo rendimiento lo que, ciertamente, abona a favor de la presunción que les atribuye una moderada validez de constructos (ver las tablas números 6, 7 y 20).

7. No hemos encontrado en la bibliografía especializada informes sobre la aplicación de pruebas de rendimiento académico a docentes de educación básica en algún país de América Latina. Por esta razón no ha sido posible comparar las estructuras de las pruebas, sus contenidos, la escala utilizada, el método de administración, el método de calificación y los indicadores estadísticos de validez y de confiabilidad. Solamente tenemos referencia en el Perú de la investigación realizada por Antonio Díaz Saucedo que en un informe de tesis (1998) reportó haber aplicado una prueba de conocimientos de matemática a una muestra de 451 estudiantes aspirantes a la docencia en dicha especialidad, del cuarto y quinto año de 11 universidades y 6 institutos superiores pedagógicos del país,

obteniendo un promedio general de 10,90 puntos, con una desviación estándar ( $\sigma$ ) de 6,51, resultados que son muy aproximados a los obtenidos por esta investigación en las pruebas del tipo D, como lo muestra la tabla n.º 7. Un resultado más lejano pero en el mismo sentido lo proporciona la noticia que aparece en el Boletín de IESALC (febrero-marzo 2005) que informa que en los exámenes de calidad de la educación superior de Colombia (Ecaes) han sido los estudiantes que están en trance de graduarse de Licenciados en Educación, con énfasis en matemática y francés, los únicos que no alcanzaron el puntaje mínimo de 120 que lograron los estudiantes de todas las demás carreras universitarias.

8. Los promedios por sedes de administración de las pruebas incluyen información sorprendente. Por ejemplo, los docentes de la ciudad amazónica de Nauta han logrado el segundo promedio en el ámbito nacional con un puntaje, en cifras redondas, sólo 39 centésimos menor que el promedio máximo obtenido por las sedes de San Borja con un valor de 14,82098. Asimismo, el promedio de Nauta tiene significación estadística porque ha sido obtenido sobre una población de 92 docentes. En este caso la relación entre rendimiento académico y pobreza parece haberse tornado en inversa porque según el mapa de la pobreza publicado por el gobierno peruano en los diarios de Lima el día 11 de febrero de 2005, Loreto tiene un 66,4% de su población en situación de pobreza y un 40,8% en extrema pobreza, mientras que la región de Lima tiene un 35,8% en situación de pobreza y un 3,8% en extrema pobreza. Sin embargo, las jurisdicciones de las unidades de gestión educativa locales de Breña, Rímac, Ate-Vitarte, San Juan de Miraflores, Comas, San Juan de Lurigancho más la Dirección Regional del Callao y la Dirección Regional de Lima Metropolitana han tenido un promedio inferior, como lo muestra la tabla n.º 9. Asimismo, el hecho de que la Dirección Regional de Loreto (Iquitos) haya ocupado el tercer lugar por una diferencia de 12 centésimos respecto de Nauta corrobora este resultado.

9. Aunque no hay grandes diferencias entre los valores de los indicadores generales, el rendimiento promedio de las damas en las 14 áreas examinadas fue mayor que el de los varones, como lo muestra la tabla n.º 6. La diferencia a favor de las damas es de 0,70 puntos que en grandes números, 59.847 damas y 34.946 varones,

es relevante. Dicha diferencia es ostensible en los casos de las pruebas de educación inicial, ciencias naturales y lengua extranjera, pues en todos los demás casos su rendimiento es muy semejante al de los varones o ligeramente menor. La excepción la constituye la prueba de matemática que fue respondida de manera significativamente mejor por los varones (13,009 *vs.* 10,038 como lo muestra la tabla n.º 8). Asimismo, la desviación estándar ( $\sigma$ ) de las damas de 6,07 es mayor que de los varones de 5,24. Esta diferencia revela que la población de varones mostró un rendimiento académico más homogéneo que el presentado por las damas.

10. Clasificando a los docentes según el tipo de institución de formación profesional en la que cursaron estudios y según el nivel educativo en el que se han especializado, se encontró que los licenciados universitarios tuvieron un rendimiento académico superior al de los titulados en los institutos superiores pedagógicos. La diferencia más acusada fue encontrada en el nivel de la educación inicial, seguida de las diferencias en el nivel de la educación secundaria y en el de la educación primaria. Los valores de la diferencia a favor de los licenciados universitarios, en cada uno de los casos anteriores, fueron de 3,2295, 2,5887 y 1,3170, respectivamente. Este hallazgo puede interpretarse como un indicador de que la formación profesional es mejor en las universidades autónomas que en los institutos superiores pedagógicos que dependen académicamente de la Dirección General de Formación y Capacitación Docente del Ministerio de Educación. Es destacable que en este caso las damas que cuentan con licenciatura en una universidad tienen, claramente, puntajes superiores a las que estudiaron en los institutos superiores pedagógicos y a los varones que también obtuvieron licenciatura universitaria, como puede observarse en la tabla n.º 13. Por ejemplo, las licenciadas en educación inicial tienen una ventaja de 3,2256 puntos sobre las tituladas en institutos superiores pedagógicos y una superioridad de 6,4740 puntos sobre sus colegas licenciados. Estos últimos se ubican a más de una desviación estándar ( $\sigma$ ) por debajo de las licenciadas, lo que equivale en puntajes Z a 1,0866. En breve, los resultados anteriores permiten deducir que la autonomía académica guarda una relación directa con la calidad de la formación docente, como lo muestra la tabla n.º 13.

11. La edad es un factor relevante en el rendimiento académico de los docentes. En la tabla n.º 17 se analizan cuatro grupos etarios: 21-25, 26-30, 31-40 y 41 a más años. Los resultados muestran que el mayor promedio de rendimiento corresponde al sector más joven (13,21 puntos) y, luego, en gradiente estricta el menor promedio corresponde al sector de mayor edad (10,931 puntos). La misma situación se reproduce, muy aproximadamente, cuando se subdivide la población en damas y varones. Una ligera alteración en la pendiente se genera en el caso de los varones en tanto que el grupo de 41 a más años pasa al tercer lugar. Este resultado entra en conflicto con las políticas educativas vigentes, hasta la fecha, que declarativamente persiguen una educación de calidad pero aplican reglamentos que concedieron en el concurso de marzo de 2002 y, posteriormente, ventajas desmedidas, en la tabla de calificaciones, a los docentes de mayor edad que son, según las evidencias disponibles, los de menor rendimiento académico.



12. Considerando que en 39 sedes se administraron las pruebas del tipo M para docentes de institutos superiores tecnológicos, se encontró que en trece de ellas los concursantes con título no pedagógico rindieron académicamente más que los graduados en institutos superiores pedagógicos, como lo muestra la tabla n.º 14. Están incluidas dentro de esta situación, entre otras, la ciudades de Lima, Piura, Puno y Tocache. Este resultado prueba que para mejorar la educación tecnológica es indispensable incorporar a las aulas de los institutos superiores tecnológicos a los profesionales técnicos que poseen vocación pedagógica aunque carezcan del denominado título pedagógico. Sin embargo, esta posibilidad de brindar un mejor servicio en el nivel de educación superior no universitaria y también en el nivel de educación básica está prohibida por la ley del profesorado, por la ley de colegio profesional de docentes y por la actual ley de educación n.º 28044. Estos dispositivos legales, por absurdo que parezca, sobreprotegen los intereses gremiales impidiendo que se nombren en los colegios públicos, por ejemplo, psicólogos, matemáticos, físicos, traductores, historiadores, etc., lo cual es común en los colegios privados de alta paga, los mismos que los prefieren motivados por la necesidad de contar con personal especializado, en disciplinas científicas, para incrementar la calidad del servicio educativo.

## Los diez problemas cardinales de la educación peruana

1. Planificación (incluyendo la fase legislativa) y ejecución inadecuadas del gasto público destinado a la educación nacional y del financiamiento complementario generado por la cooperación internacional debido a que existe un severo desfase entre los órganos capaces de aportar información científica sobre las necesidades sociales de educación y desarrollo y las políticas educativas decididas y ejecutadas por el Ministerio de Educación y por los órganos descentralizados del Estado.
2. Ministerio de Educación con una textura burocrática y clientelista ajena al mundo académico, desconectada de las universidades, de los colegios profesionales, de las sociedades científicas y de los centros internacionales de producción de ciencia y tecnología.
3. Dirección centralista, anticientífica, autoritaria, monocorde y rígida de los currículos de la educación básica, los mismos que se encuentran desconectados de la construcción de un proyecto nacional de desarrollo económico con equidad que se sustente en el conocimiento científico y tecnológico disponible en el mundo globalizado.
4. Conducción centralista, anticientífica, autoritaria, monocorde, rígida y sobredimensionada de la formación docente a través de más de 390 Institutos Superiores Pedagógicos (ISP). A esto se suma un sistema universitario de formación docente a través de facultades de Educación con estructura obsoleta y desarticulada

de las facultades no-pedagógicas. Asimismo, sistema de capacitación docente oneroso, centrado en el *cómo* de la educación y de espaldas al *qué*, lo que refuerza la tendencia anticientífica y oscurantista.

5. Rendimientos escolares en matemática y en lenguaje, en los niveles de educación primaria y de educación secundaria, extremadamente deficientes. De acuerdo con las pruebas de la UNESCO (1998-2000) nos disputamos con Haití el último lugar del *ranking* de la región de América Latina y del Caribe. Y según las pruebas PISA-2000, administradas por la OECD, ocupamos el último lugar en una muestra de 41 países.
6. Ausencia de un instituto o centro especializado de investigación educacional y de un sistema de pruebas nacionales, especialmente para escolares de 11 y 16 años y para egresantes del sistema universitario.
7. Carrera magisterial corporativista sustentada en una legislación pedagógica que excluye del sistema de educación básica, entre otros, a los profesionales especializados en el aprendizaje humano, la salud y la gestión, potenciando desmedidamente a los recursos menos especializados. Se incurre en el anacronismo de reducir el concepto de profesional de la educación al de docente de aula. Adicionalmente, el ingreso al servicio no es por concurso público a través de un sistema de exámenes que evalúen la suficiencia profesional.
8. Rendimientos académicos de los docentes de educación básica altamente deficitarios en las materias científicas y humanísticas cuyo aprendizaje deben dirigir en el aula. Igualmente, en los conocimientos psicopedagógicos que requiere su desempeño. Los resultados de las pruebas de suficiencia profesional administradas en marzo del 2002 proporcionan indicadores estadísticos que guardan correspondencia estrecha con las evaluaciones internacionales de los alumnos. En breve, se puede afirmar que los pobres rendimientos mostrados por los escolares peruanos en las pruebas UNESCO y PISA constituyen un fiel reflejo de la deficiente formación académica de sus profesores.

- 
- 
9. Universidades, en general, monadizadas, estructuradas legalmente bajo el desfasado modelo de Córdoba (1918) como organizaciones político-partidarias y no científicas, desconectas entre sí y marginalmente posicionadas dentro de la comunidad internacional. Subsisten sin nexo con el sistema productivo como expedidoras de certificaciones para el mercado de servicios profesionales, al margen de las necesidades del desarrollo social, desarticuladas del sistema de educación básica y encapsuladas en sus intereses y conflictos internos.
  10. Universidades, en gran medida, formadoras de profesionales de servicios y no de producción, encubridoras de la desocupación y del subempleo, sin liderazgo social, y sin aporte tangible y científicamente fundado al desarrollo económico y a la reducción de la inequidad. Bajo estas condiciones operan como órganos catalizadores de los conflictos y de la violencia sistémica en permanente riesgo de ser instrumentalizados y desregulados.

