

Las Imprecisiones Evaluativas en la Corrección Escrita Debido a la Falta de Criterios Definidos

▮ Armando Terribili Filho *

▮ Andreia Melanda Chirinea **

Resumen

La presente es una investigación de campo realizada con diez profesores que son estudiantes de un curso *stricto sensu* de posgrado en Educación, y pretende exponer la importancia de tener criterios bien definidos al realizar la evaluación de pruebas. La investigación se basa en la simulación de una prueba hecha por dos estudiantes hipotéticos de la enseñanza fundamental, que fue corregida individualmente por los profesores. La primera corrección se hizo con criterios individuales; para la segunda corrección, el grupo de profesores discutió y estableció los criterios que serían usados, después corrigió nuevamente de forma individual. Los resultados muestran mejoras significativas en la evaluación, ya que la distorsión de la clasificación se redujo. Sin embargo, los resultados indican que la evaluación aún requiere una reflexión crítica por parte de los docentes y coordinadores educativos, no sólo al corregir las pruebas, sino también en el proceso de evaluación y sus instrumentos, elaboración de planteos, lo que señala la importancia del binomio "¿qué medir?" y "¿cómo medir?".

Palabras clave: Evaluación. Corrección de pruebas. Criterios de corrección.

* Postdoctor y doctor en Educación por la UNESP de Marília (SP), máster en Administración. Profesor de la Universidad Positivo y de la ESIC de Curitiba (PR). E-mail: armando@impriamo.com.br.

** Máster y cursando el doctorado en Educación por la UNESP de Marília (SP). Profesora de la red pública municipal de Bauru (SP) y profesora de la Universidad *Sagrado Coração*. E-mail: andreia.melanda@gmail.com.

1. Introdução

Brasil posee actualmente más de 46 millones de estudiantes. Según datos del Censo de Educación Primaria de 2014 (INEP, 2014a), la educación infantil (jardín de infantes y preescola) tiene 5.492.301 niños inscriptos; la enseñanza primaria, 23.446.790 y el secundario 7.031.624 estudiantes. Se debe agregar a estos números 2.985.304 de inscriptos en el proyecto de Educación de Jóvenes y Adultos (EJA), siendo 68% en la enseñanza primaria y 32% en la enseñanza media. En la educación superior, según los datos del Censo de la Educación Superior de 2013, hay 7.526.681 inscripciones en la Educación Superior - cursos de grado, secuenciales y posgrado *stricto sensu*. (INEP, 2014b).

Excepto los alumnos de la educación infantil, los 41 millones de la Primaria, Secundaria y Superior realizan procesos evaluativos en el aula, para identificar la efectividad del aprendizaje. Claro que los procesos evaluativos deben realizarse a través de distintos instrumentos para que el alumno pueda ser evaluado de forma holística, incluyendo: pruebas escritas, trabajos de investigación, seminarios, estudios de caso, paneles, dinámicas de grupo, dramatización, entre otros.

Sin embargo, lo que se constata es que la mayor parte del sistema evaluativo lo constituyen pruebas escritas y trabajos, con lo cual se calcula el promedio final del alumno, y, basado en éste (en general, con peso mayor para la prueba escrita), se califica al alumno, siendo entonces, aprobado o reprobado.

La prueba escrita, además de ser “el gran instrumento” de evaluación presenta bastantes problemas, especialmente para la forma de medir el aprendizaje de los estudiantes, que tiene como característica el proceso “devolutivo” del alumno de aquello que le fue enseñado, es decir, la valoración de la reproducción, memorización, lo que limita el análisis y estimula poco la capacidad de deducción, inducción, creación, realización de análisis sistemático, síntesis y, finalmente, reflexión crítica. De cualquier manera, el objetivo de este artículo no es discutir los aspectos conceptuales de la prueba escrita, sino mostrar cuán frágil es el proceso de corrección de las pruebas.

Una prueba puede presentar planteos mal formulados que originan dudas de interpretación o incluso contienen errores. También, el estado físico y mental del profesor en el momento de la corrección puede influir en el juicio de valor que dará. Peor todavía,

los criterios personales y subjetivos de evaluación del profesor influyen en el resultado de la evaluación.

Para poder diagnosticar si el proceso de corrección de pruebas torna vulnerable la evaluación del estudiante, se realizó un estudio de campo con la elaboración y corrección de una prueba hipotética, con planteos a nivel de cuarto grado primario, con los siguientes temas: lengua portuguesa, geografía e historia, cada planteo o pregunta valiendo 2,5. Se simuló que dos alumnos habían hecho la prueba (María Clara y Juan Pedro). Esas pruebas las corrigieron 10 profesores que son alumnos de un curso de posgrado *strictus sensu* en educación de una institución pública ubicada en una ciudad del interior paulista. Para la prueba de María Clara, un profesor le puso 2,0 y otro, 7,5, lo que representó una variación de 5,5 puntos! Para la prueba de Juan Pedro, la nota mínima fue 3,0 y la máxima: 9,0, lo que significó una variación mayor, ¡6,0 puntos! No se puede ignorar que esa discrepancia es preocupante, pues indica que los juicios efectuados están muy alejados; y algo peor, esos valores mínimos de 2,0 y 3,0 y los máximos de 7,5 y 9,0 significan reprobación y aprobación de un alumno.

Para la segunda corrección, para las mismas pruebas se estipularon criterios claros y objetivos: (1) la respuesta del alumno debe dirigirse al planteo propuesto; (2) la clave de respuesta no es binaria; (3) errores de portugués deben ser considerados en la corrección; (4) aspectos personales del alumno deben ser ignorados; y (5) no se debe efectuar ningún tipo de redondeo.

En la corrección efectuada, la variación de la prueba de María Clara disminuyó, pero, todavía quedó en nivel inadecuado de 4,0 puntos. Lo mismo ocurrió con la prueba de Juan Pedro, con variación de 4,65.

Así, alumnos del primario realizan aproximadamente 6 pruebas mensuales: los de la enseñanza Media, 10 pruebas bimestrales y los de enseñanza Superior, 20 pruebas semestrales, de lo que se puede deducir que hay un universo de injusticias en cerca de 2 billones de pruebas escritas realizadas y corregidas anualmente en el país.

Cuando se reflexione sobre los cambios de enfoque en la evaluación, la recogida de los datos y la sistematización de los mismos será posible explicitar mejor como la evaluación se constituye en juicio de valor para el “bien” o para el “mal” dependiendo de los criterios preestablecidos por el evaluador.

2. Evaluación: caminos y sentidos

La evaluación escolar es una de las etapas más importantes del proceso de enseñanza-aprendizaje y uno de los mecanismos de sustentación de la lógica organizacional del trabajo y de la cultura escolar. Suhr (2008) destaca que entre las habilidades necesarias para la competencia docente está la de evaluar adecuadamente el nivel de aprendizaje de los estudiantes, para favorecer un aprendizaje significativo.

Por medio de la evaluación el profesor consigue replanificar su trabajo para adecuar mejor los objetivos del curso y del aprendizaje. Para tal el profesor necesita atribuir un juicio de valor a cada etapa laboral ejecutada por el estudiante. Aunque sea un instrumento importante para el trabajo del profesor, dependiendo del juicio de valor atribuido y del tipo de evaluación realizado, puede legitimar el éxito o el fracaso escolar.

Hadji (1994) reconoce la dificultad de realizar una evaluación principalmente por el juicio de valor que se le atribuye al trabajo que el alumno desarrolla y que el profesor evalúa. El principal planteo del autor es la ambigüedad del valor atribuido, así como la definición de las variables que serán evaluadas.

El mayor escollo para la evaluación educacional es la subjetividad presente en el acto de evaluar y, consecuentemente, de atribuir una nota o un valor para un determinado trabajo realizado por el alumno. En tal sentido cabe preguntar: ¿Cómo evaluar un trabajo que otra persona hizo y qué medida debe usarse para atribuirle un valor, una nota o un concepto?

Para el autor es necesario alejarse del sujeto evaluador y del objeto evaluado. Esta distancia es importante para construir fundamentos concretos para atribuir valor que sólo están más allá de la subjetividad. Para concretar esta perspectiva de evaluación es necesario establecer criterios para evaluar, es decir, definir claramente qué y cómo evaluar. (HADJI, 1994).

HADJI (1994) considera que estos criterios necesitan ser claros y precisos, bajo la perspectiva de que el evaluador considere el proceso del “venir a ser”, es decir, considerar en este proceso lo real y lo ideal. Y solo criterios puntuales definidos preliminarmente son capaces de dirigir la mirada del evaluador para puntos significativos. Entonces, vemos que hay una articulación entre la realidad, o la situación concreta observada, y lo que se espera idealmente. Bajo este aspecto Hadji (1994) considera que:

Articulación, en primer lugar, entre lo referido y lo referente, ya que evaluar consiste en atribuir un “valor” [...] a una situación real a la luz de una situación deseada, al confrontar así el campo de la realidad concreta con el de las expectativas. (HADJI, 1994, p. 32).

Si partimos del principio que evaluar es un proceso que proyecta acciones reales para acciones ideales, entonces, la evaluación presupone una relación entre lo que existe y fue diagnosticado y lo que era esperado, es decir, la consecución de los objetivos educacionales.

Durante mucho tiempo la evaluación escolar tuvo como presuposición el positivismo, y su finalidad era comprobar y comparar comportamientos. Según Sousa (1998), muchos teóricos como Tyler, Bloom, Gagné, Mager y Popham consideraban la evaluación un proceso de medir objetivos. Sousa (1998) afirma que para estos autores el proceso evaluativo verificaba cuánto un individuo consigue alcanzar de los objetivos educacionales previamente establecidos. Midiendo los objetivos se podría también medir los patrones de comportamiento deseados en el alumno, ya que la evaluación es el principal instrumento destinado a verificar los cambios comportamentales, así como el grado en que ellos ocurren.

En los años de la década del 80, diversos autores que analizan el tema cuestionan esta vertiente psicológica de evaluación tratando de desarrollar un estudio crítico de la evaluación ejercida en la vida cotidiana escolar. Dicho estudio preconiza que la evaluación es una actividad socialmente determinada, sea por la función de la escuela, por la propia sociedad o por el tipo de hombre que se quiere formar. En estos casos los determinantes sociales definen la función de la escuela y el instrumento para legitimar esta función acaba siendo la propia evaluación. (SOUZA, 1998).

Esta influencia trae la necesidad de construir instrumentos evaluativos capaces no sólo de verificar el desempeño escolar, sino, principalmente, comprender los procesos de construcción de las desigualdades sociales en el interior de la escuela, con lo cual:

La toma de conciencia, por parte del profesor, de los efectos del nivel socioeconómico de los alumnos sobre los resultados de aprendizaje no implica la realización de evaluaciones que ocultan dicho hecho. Es necesario identificar dónde el alumno presenta mayores dificultades, cómo se muestran estos desfases y garantizar una escuela y un sistema de enseñanza que realmente ofrezca educación con justicia. (SOUZA, 1998, p. 168).

Fuera estos planteos, otro argumento usado por la autora es el propio cuestionamiento que el profesor debe hacerse preguntándose sobre su práctica evaluativa, entender que ese proceso conlleva mucha inseguridad, miedos, ansiedad, deseos y que, a menudo, estos sentimientos interfieren en el proceso de evaluar.

Luckesi (1998) señala algunas características para el proceso de evaluación y considera que en un primer momento la evaluación se constituye en un juicio de valor sobre determinada realidad usando criterios preestablecidos por el evaluador. El valor juzgado va a considerar la satisfacción de tales criterios, así como la realidad, que es el objeto de la evaluación.

Otra característica señalada por Luckesi (1998) es el sentido del juicio para una toma de decisión. En este sentido el evaluador podrá reconducir el proceso para perfeccionar los indicadores.

Según el autor, estas características acaban por impregnar procesos arbitrarios de la autoridad pedagógica, debido a que en la evaluación efectuada por la mayoría de las escuelas el foco incide sobre la calificación y no en el diagnóstico. Así, la arbitrariedad está en determinar posiciones de los estudiantes indicando conceptos como insuficiente, suficiente o plenamente suficiente. Esta calificación impide que la evaluación cumpla con la función de reorientar la práctica pedagógica de forma más crítica, menos subjetiva y más cuidadosa en lo que se refiere al juicio de valor y en elegir lo que realmente es importante juzgar.

Penna Firme (1994) dice que, a menudo, la evaluación trae consecuencias perversas e inútiles para el proceso educativo y que se debe entender todo el movimiento realizado durante cien años de estudios sobre el tema, buscando entender los avances y los retrocesos sobre la evaluación.

Tanto Penna Firme (1994) como Fernandes (2009) utilizan las categorías de Guba y Lincoln (1989) al describir la trayectoria de los estudios sobre la evaluación educacional. Según los autores, en un primer momento, entre las décadas de 1920 y 1930, no había distinción entre evaluación y medida, y los estudios sobre evaluación educacional se dirigían a instrumentos y test que pudieran verificar el rendimiento escolar. La evaluación durante este período era vista como algo eminentemente técnico y su función era esencialmente clasificar y calificar a los estudiantes para después verificar su desempeño.

La segunda generación de estudios sobre evaluación educacional (1930-1940) fue denominada descriptiva porque buscaba describir lo que sería éxito o fracaso según los objetivos establecidos. La evaluación todavía continuaba a ser vista como algo técnico. Sin embargo, avanzó en lo que se refiere a la actuación del evaluador que usa patrones y criterios basados en el establecimiento de objetivos educacionales. Se observa que a medida que los estudios sobre la evaluación se van consustanciando de una generación a otra, más se avanza en el entendimiento y en la práctica de lo que viene a ser la evaluación educacional, por lo menos teóricamente.

También, según los autores, la tercera generación de estudios sobre evaluación educacional se caracterizaba por las mismas limitaciones de la etapa anterior, principalmente la falta de definición y claridad de los objetivos educacionales. Para poder solucionar esta problemática, según Penna Firme (1994) y Fernandes (2009), los especialistas Stake y Screven consideraron necesario establecer el juicio de valor, principal característica de la tercera generación. De este modo, el evaluador tendría el poder de juzgar, teniendo como referencia tanto la medida como la descripción, fruto de los estudios de las generaciones anteriores. Según Penna Firme (1994, p. 108): “no bastaría medir y describir, era necesario juzgar sobre el conjunto de todas las dimensiones del objeto, incluso sobre los propios objetivos”. Esta generación estimuló varios modelos de evaluación para concebir un juicio de valor más claro y eficiente.

La cuarta y última generación, ya en la década de 1990, al identificar los problemas presentados por las generaciones anteriores, buscó basar sus estudios en la negociación de la evaluación. Bajo esta perspectiva diferentes autores buscan un consenso de forma negociada, interactiva y pautada bajo el paradigma del constructivismo. En este escenario, Penna Firme (1994) refuerza que,

Siguiendo esta perspectiva, cuánto mayor es la participación al elegir las preocupaciones, los planteos evaluativos y los métodos y la interpretación de los resultados, mayor es el nivel de negociación y más profunda es la capacidad de responder a los interesados en el proceso y en el objeto de la evaluación. Es, pues, aspecto crucial de tal participación, la identificación de esos interesados. (PENNA FIRME, 1994, p. 109).

Por lo tanto, la cuarta generación entiende que para construir una evaluación eficaz hay que captar variables que van más allá de las prerrogativas de control, pues este

proceso conlleva aspectos culturales, sociales, políticos y éticos. De lo cual deriva la necesidad de negociación entre observador y observado. Aunque la cuarta generación de estudios sobre evaluación educacional haya contemplado la evaluación negociada, todavía hoy predominan prácticas evaluativas que tienen por objeto la calificación, selección y certificación de los estudiantes, en detrimento de prácticas que buscan comprender mejor las competencias adquiridas por los alumnos de forma más clara.

2.1. La prueba escrita

Camargo (2011) reunió trabajos desarrollados por especialistas en Educación para discutir el tema “evaluación escrita”, que todavía es el principal instrumento de aferición en la escuela, con el siguiente título “Profesor puesto a prueba”. De este artículo basado en libros, resultados de investigaciones y artículos de Cipriano Luckesi, Dirce Aparecida Foletto de Moraes, Vasco Moretto, Bernadete Gatti, Tadeu da Ponte y Jeffrey Karpicke, se pueden extraer diez elementos importantes de la evaluación.

2.1.1. Origen de la prueba

Según Cipriano Luckesi, la prueba surgió en Europa en el siglo XVI como un recurso de recolección de datos sobre el desempeño de alumnos. Actualmente, la prueba es el más común instrumento usado en el aula, pues permite reunir informaciones sobre una gran cantidad de alumnos.

2.1.2. Instrumento de documentación y comprobación

Para Dirce Aparecida Foletto de Moraes, la prueba documenta y registra el conocimiento, permite la representación final a través de un valor numérico, el cual retrata el aprendizaje del alumno evaluado. La profesora afirma que “los educadores incluso buscan nuevos caminos, pero, por desconocer las diferentes herramientas o por sentirse inseguros, se garantizan en la prueba como instrumento comprobatorio”.

2.1.3. Comunicación familia, escuela y chicos

La prueba es un histórico y reconocido medio de comunicación entre familias, alumnos y escuela. Dirce Aparecida Foletto de Moraes confirma tal situación al afirmar que es más frecuente un padre preguntarle al hijo “¿cómo te fue en la prueba?” que “¿qué aprendiste?” Además, la prueba se tornó un instrumento de estatus entre alumnos, induciendo la existencia de una jerarquía en el aula.

2.1.4. La prueba no es algo malo

Si hay críticas sobre la prueba por su uso inadecuado (control de la clase y del poder del profesor sobre los alumnos), Camargo (2011) afirma que ella debe ser adecuadamente aplicada, ejemplificando que una prueba escrita es útil cuando se limita a las competencias lógico-verbales.

2.1.5. Elaboración de la prueba y parámetros de corrección

Tadeu da Ponte destaca la importancia de la claridad sobre la elaboración de cada planteo, la objetividad y la adecuación del vocabulario según la edad de los alumnos. Para Vasco Moretto, una característica típica de una buena prueba es su capacidad de estimular el aprendizaje significativo, lo que requiere la contextualización de lo que se pregunta. Vasco Moretto critica la falta de parámetros claros para la corrección.

2.1.6. Instrumento de reflexión para profesores

Dirce Aparecida Foletto de Moraes afirma que "el *feedback* permite que él [el profesor] se sitúe con relación a los aprendizajes, tenga condiciones de entender y no sólo constatar sus errores y aciertos. Hace falta reflexionar [...]".

2.1.7. “Aprender de memoria” puede ser importante

Una investigación dirigida por el psicólogo Jeffrey Karpicke de la Universidad de *Purdue* (Indiana, EEUU) con 200 jóvenes universitarios señaló que el "aprender de memoria" puede impactar positivamente el desempeño de los estudiantes. Según el estudio, la memorización ayudó a los alumnos a responder planteos que exigían deducciones más complejas y cruzamiento de informaciones, lo que obligaría al cerebro a reorganizar el

asunto y priorizar determinados tópicos. Así, la memorización, considerada un método anticuado por muchos educadores, no puede ser abandonada. El comentario de Tadeu da Ponte sobre el asunto es que "existen, de hecho, conocimientos que precisan ser memorizados. Por otro lado, debemos buscar ofrecer situaciones-problema que lleven al alumno a apropiarse de esos contenidos, haciéndolo capaz de pensar de forma diferente, según los contextos que se presentan".

2.1.8. Profesor repite el modelo

Camargo (2011) comenta que Gatti y Barreto (2009), cuando analizaron los currículos de cursos de pedagogía de todo el país, identificaron que, en general, el docente reproduce prácticas que vivió en su curso de grado. En las facultades, el método más común de evaluación es la prueba escrita, por eso, posteriormente la aplicará a sus alumnos. El problema es que las consecuencias de esa falta de preparación remiten directamente a planteos más graves de la educación brasileña, como el fracaso escolar.

2.1.9. La prueba – claridad, objetividad y contexto

La prueba, según la síntesis de Camargo (2011), debe ser planeada con antelación, bien elaborada y bien presentada. Cada pregunta debe ser clara, con objetividad e intencionalidad, adecuadas al nivel de los alumnos, coherentes con las clases impartidas por el profesor y contextualizadas para provocar el razonamiento del alumno.

2.1.10. La prueba – ¿qué evitar?

Según Camargo (2011), deben evitarse pruebas exhaustivas y planteos con "trampas", planteos con uso de negativas, que exijan solamente memorización o que aborden los extremos de los contenidos estudiados, es decir, no debe ser muy general, ni muy específica, solicitando la información contenida en una "nota a pie de página".

Suhr (2008) presenta nuevas posibilidades sobre la elaboración de pruebas en lo que se refiere a planteos objetivos y disertativos. Entre los planteos objetivos, presenta los de respuesta corta, de completar, de correcto e incorrecto, de correlación o combinación y de *multiple-choice*. Para los planteos disertativos, el autor subraya que el "verbo" puede

definir lo que se pretende del alumno. Por ejemplo: relacionar o enumerar (exige sólo que el alumno recuerde), organizar (solicitar aspectos de memorización, pero con criterio cronológico o de importancia), seleccionar (solicita elección basada en normas de apreciación), describir (solicita la exposición de características de un hecho, objeto o fenómeno), analizar (pide la estructuración de argumentos, relacionándolos), definir (solicita que se enuncien los atributos esenciales y específicos de un hecho, objeto o fenómeno), ejemplificar (exige la aplicación causa y efecto), comparar (exige el análisis simultáneo de hechos, objetos o fenómenos), sintetizar (pide la exposición de forma concisa de una idea o asunto), esquematizar (pide la elaboración de síntesis organizada en tópicos y subtópicos), interpretar (solicita el análisis del significado de algo) y criticar (exige juicio y elaboración de análisis crítico basado en patrones o criterios).

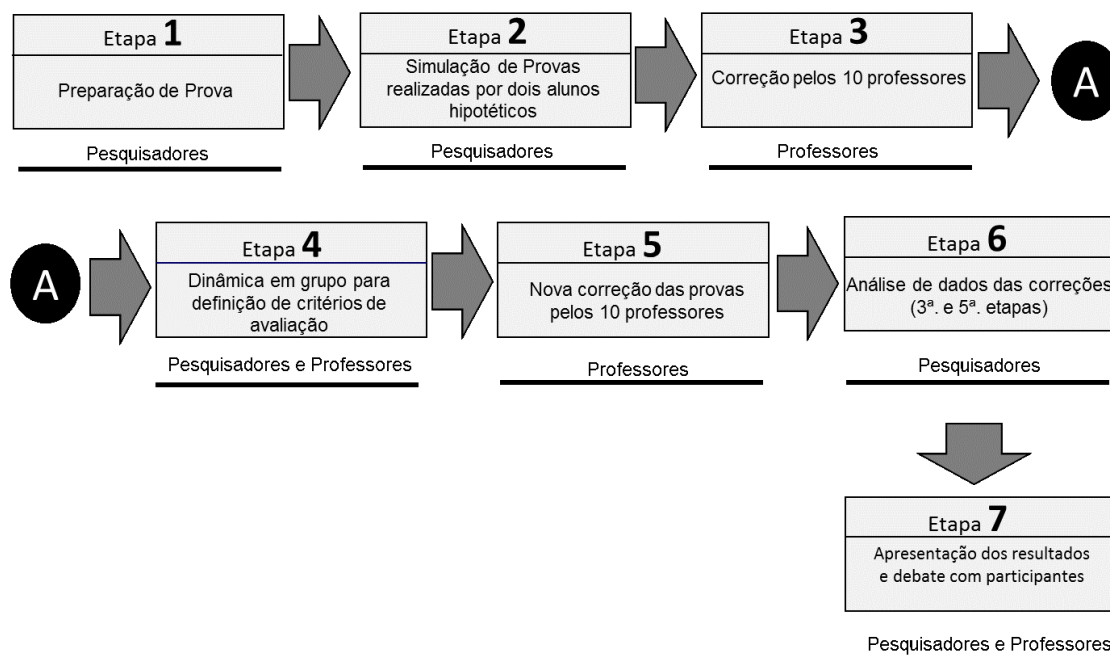
3. Recolección de datos

Para demostrar que la evaluación es un juicio de valor personal en detrimento de prácticas que buscan comprender de forma clara las competencias adquiridas por los alumnos se realizó una investigación en el aula con diez alumnos de un curso de posgrado que actúan profesionalmente como profesores.

Se trata de profesores entre 25 y 40 años, que actúan hace por lo menos tres años en la enseñanza Primaria o Media, con predominio en los grados iniciales de la enseñanza Primaria (niños de 6 a 10 años).

Así, después de la revisión bibliográfica del tema “evaluación”, la investigación de campo fue dividida en siete etapas distintas y continuas, como lo muestra la Figura 1. En ella, cada etapa se representa a través de un rectángulo debidamente identificado, y bajo de cada rectángulo se presentan los autores de la referida etapa.

Figura 1 – Procedimientos metodológicos



Fuente: los autores (2014).

En la primera etapa se elaboró una prueba con cuatro planteos abiertos, de nivel relativo al cuarto grado primario, tratando los siguientes temas: lengua portuguesa, matemáticas, geografía e historia. Los planteos de la prueba, aunque sencillos y con bajo nivel de complejidad, permitían intencionalmente más de una respuesta correcta, aunque no fueran convencionales. Los planteos pretendían solamente verificar el rendimiento escolar, que, según Penna Firme (1994), era una característica dominante de la primera generación de la evaluación (década de 1920). En el Cuadro 1, los cuatro planteos de la prueba.

Cuadro 1 – Contenido de la prueba

Planteo 1 – Utilice las palabras: viaje, capital, hotel, paseo y cultura, escriba sobre un viaje hecho al Estado de Santa Catarina.

Planteo 2 – Hoy es el cumpleaños de Daniela y Sergio. Él cumple 13 años y ella el doble de él. ¿Cuántas velas debo comprar para hacerles la fiesta a ellos?

Planteo 3 – Don Pedro II fue emperador de Brasil por casi 50 años durante el siglo XIX. ¿Qué ciudad fundó en el Estado de Río de Janeiro y cuándo terminó el imperio en Brasil?

Planteo 4 – Señale en el mapa la capital del país y el Río Amazonas.



Fuente: los autores (2014).

En el Planteo 1, el alumno podría crear una frase bien estructurada, y equivocarse sobre el nombre de la Provincia de Santa Catarina u omitir el nombre de la ciudad. Podría mencionar que la cultura predominante en la región es la alemana, aunque tal conocimiento no se le hubiera preguntado. La formulación del planteo se opone a la perspectiva de la dificultad de realizarse una evaluación registrada por Hadji (1994) sobre la definición de las variables que serán evaluadas.

El planteo 2 estimula varias interpretaciones, pues fue mal formulado. Sergio con 13 años y Daniela con 26, podría comprar: (i) 39 velitas, siendo 13 para Sergio y 26 para Daniela, (ii) 26 velitas, cantidad suficiente para ambos festejar de forma aislada, (iii) 4 velitas, con los números 1,3,2 y 6, (iv) dos velitas, una para cada uno, o (v) sólo una velita, pues podría tenerse una única torta (o pastel) para los dos cumpleaños conteniendo una vela única para ambos.

El planteo 3 es más objetivo, pero exige del alumno sólo memorización; de cualquier modo, el término “cuándo fue el fin del Imperio” puede inducir a diferentes respuestas (año específico, período o siglo).

Finalmente, el Planteo 4 también origina confusión. Sobre la posición del Río Amazonas en el mapa de Brasil, el alumno puede poner una “x” en el lugar aproximado, o hacer un gran círculo en la región, impidiendo una evaluación más precisa. Según la concepción de Hadji (1994) se debe definir claramente qué se va a evaluar o cómo hacerlo; en este caso, se trataba de un planteo sin esos propósitos.

En la segunda etapa, los investigadores simulaban la realización de la prueba por parte de dos alumnos (María Clara y Juan Pedro) y agregaron datos personales hipotéticos de esos dos alumnos, como: edad, nivel socio-económico, relaciones familiares y condición en la asignatura en términos de necesidad de nota en la mencionada prueba para aprobación. Estos factores se definieron para posibilitar que evaluaciones basadas en valores personales de los profesores pudieran surgir e influir en la corrección.

En el tercer planteo se presentaron dos copias de las dos pruebas. Los diez participantes de la investigación recibieron una copia de cada prueba para ser corregida. Se debe subrayar que no se orientó a los participantes sobre cómo corregir la prueba, ni se les informó sobre los criterios (correcto, incorrecto, más o menos correcto), ni las particularidades (notas con o sin redondeo), ni el proceso de atribución de valor para cada planteo ni tampoco los pasos siguientes de la investigación realizada. En un primer momento, algunos participantes se fastidiaron con la propuesta de los investigadores reforzando lo dicho por Souza (1998) cuando afirma que la evaluación es el principal instrumento para verificar los cambios comportamentales del alumno y el grado con que ocurren, y la investigación no se proponía efectuar esta validación (y no tenía medios para ello), sólo calificar las respuestas dadas y cuantificar las puntuaciones. Sin embargo, por ser profesores, aceptaron realizar la tarea y colaborar con la investigación en curso, al considerar que “la docencia empieza por la investigación [...], la clase es consecuencia de la investigación y se fortalece por la investigación”. (DEMO, 2012, p. 86-87).

Cuando terminaron de hacer la prueba, se recogieron los materiales. La etapa siguiente fue la realización de una dinámica con el grupo para definir los criterios de corrección de las pruebas, actividad coordinada por los investigadores junto a los

participantes. Se definieron, entonces, cinco criterios, intencionalmente, no englobando la clave de respuestas de la prueba o algo parecido, para mantener la personalidad de “juicio de valor”. Los cinco criterios definidos fueron:

- (1) la respuesta debe dirigirse a la propuesta del planteo de la pregunta (aunque sea parcial);
- (2) la clave de respuestas es “amplia” (no binaria);
- (3) errores de portugués deben reducir la nota (hasta 20% de cada planteo);
- (4) aspectos personales del alumno no pueden ser considerados; y
- (5) no efectuar ningún tipo de redondeo.

En la quinta etapa los participantes efectuaron una segunda corrección de las pruebas, basados en criterios definidos. Después de recoger los materiales, hubo un breve debate sobre las percepciones de cada uno en cada momento de la investigación.

La sexta etapa, ocurrida *a posteriori*, fue el análisis de los datos recogidos, la identificación de los valores mínimos y máximos obtenidos en cada prueba, el cálculo de los promedios y desvíos patrón para que se evaluaran los resultados obtenidos en la 3ª y 5ª Etapas del estudio (primera y segunda corrección). Los resultados se presentaron y se debatieron con los participantes de la investigación en la séptima y última etapa de dicha investigación.

4. Análisis de Datos

Como se dijo anteriormente, para el desarrollo de la investigación se simuló la realización de dos pruebas por parte de dos alumnos: María Clara y Juan Pedro. Las pruebas tenían preguntas de portugués, matemáticas, historia y geografía.

La primera corrección de las dos pruebas ocurrió basada en criterios personales de cada uno de los profesores analizados, sin el debate previo sobre criterios y patrones. Después que se realizó la corrección de las pruebas, se las recogió.

Enseguida, se definieron los criterios, basados en el debate entre los analizados que llegaron a la conclusión de que: lo que debería analizarse es si las respuestas dadas por los alumnos se dirigen a la propuesta de los planteos: además, incluso sin discutir las posibles respuestas correctas, el grupo decidió que deberían aceptarse las respuestas

(con sentido correcto), aunque estuvieran fuera del patrón convencional (clave de respuestas). Algunos ítems operacionales fueron: que los errores de lengua portuguesa podrían reducir hasta 20% el valor atribuido a cada planteo, que no habría ningún tipo de redondeo (sea a nivel del planteo o del valor final de la prueba), y finalmente, que los aspectos personales del alumno no podrían considerarse en la evaluación.

4.1. La prueba de María Clara

En la prueba, constaban algunos datos personales de la alumna, como: edad y nivel socioeconómico (10 años de edad; clase media), según el Anexo A.

4.1.1. Primera corrección

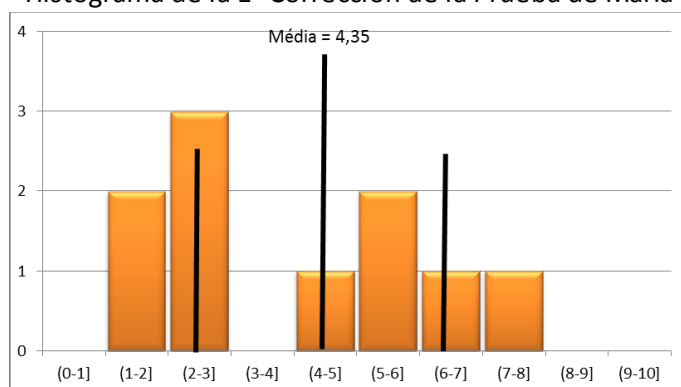
Cuadro 2 – Resumen de la 1ª corrección de la Prueba de María Clara

Nota Mínima	2,0
Nota Máxima	7,5
Nota Promedio	4,35
Variación entre Nota Mínima y Máxima	5,5
Desvío Patrón	2,01

Fuente: Los autores (2014).

Se debe subrayar que para una misma prueba, un profesor le atribuyó nota 2,0 (dos), mientras que otro le atribuyó nota 7,5 (siete y medio), representando una variación de 5,5 puntos. En la Figura 2 se puede observar el histograma de las notas atribuidas en la primera corrección para la prueba de María Clara, incluyendo la representación gráfica del promedio y el desvío patrón (dispersión en torno del promedio).

Figura 2 – Histograma de la 1ª Corrección de la Prueba de María Clara



Fuente: Los autores (2014).

4.1.2. Segunda corrección

Después de definirse en conjunto los criterios de corrección de la prueba, los diez mismos profesores corrigieron la prueba de María Clara, con los siguientes resultados, a seguir en el Cuadro 3.

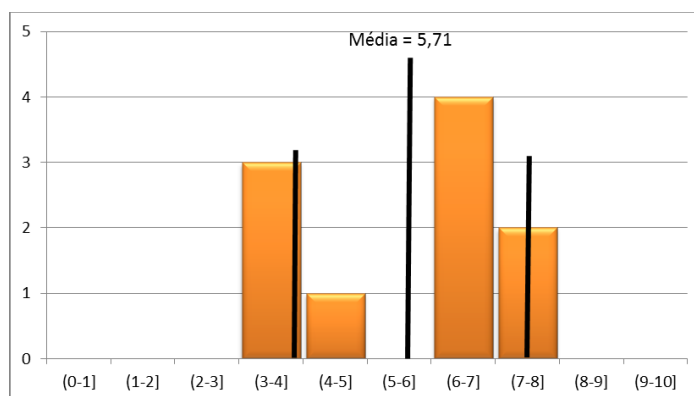
Cuadro 3 – Resumen de la 2ª corrección de la Prueba de María Clara

Nota Mínima	3,5
Nota Máxima	7,5
Nota Promedio	5,71
Variación entre Nota Mínima y Máxima	4,0
Desvío Patrón	1,74

Fuente: Los autores (2014).

Se debe subrayar que para una misma prueba, un profesor le puso nota 3,5 (tres y medio), mientras que otro, nota 7,5 (siete y medio), representando una variación de 4,0 puntos. En la Figura 3 se presenta el histograma de las notas atribuidas en la segunda corrección para la prueba de María Clara, incluyendo la representación gráfica del promedio y el desvío patrón (dispersión en torno del promedio).

Figura 3 – Histograma de la 2ª Corrección de la Prueba de María Clara



Fuente: Los autores (2014).

4.1.3. Comparativo – Prueba de María Clara

En el Cuadro 4 se presenta una síntesis comparativa entre la primera y la segunda corrección de la prueba de María Clara

Cuadro 4 – Resumen de las correcciones de la Prueba de María Clara

ítem	1ª Corrección	2ª Corrección	Variación	Comentarios
Nota Mínima	2,0	3,5	1,5	Nota mínima subió de 2,0 para 3,5.
Nota Máxima	7,5	7,5	0,0	Nota máxima no cambió.
Nota Promedio	4,35	5,71	1,36	El promedio aumentó de 4,35 para 5,71.
Variación entre Nota Mínima y Máxima	5,5	4,0	-1,5	La variación entre la nota mínima y máxima cayó de 5,5 para 4,0.
Desvío Patrón	2,01	1,74	-0,27	El desvío patrón (dispersión en torno del promedio) disminuyó de 2,01 para 1,74.

Fuente: Los autores (2014).

Como se puede observar, la segunda corrección fue más consistente que la primera. Del punto de vista matemático, queda claro por la variabilidad de los datos, que:

- 1) La variación entre la nota mínima y máxima (amplitud) se redujo en 1,5, pasando de 5,5 para 4,0; y
- 2) El desvío patrón, que mide la dispersión en torno del promedio, se redujo en 0,27, pasando de 2,01 para 1,74.

Además, como beneficios complementares que se obtuvieron con la homogeneización del proceso de corrección, se obtuvo un aumento del promedio de la

nota en 1,36, pasando de 4,35 para 5,71. Otra información importante es que si se considerara la nota 5,0 (cinco) como siendo la mínima para aprobar, la alumna María Clara, por la primera corrección realizada por los profesores, sería aprobada por sólo cuatro de ellos; por la segunda corrección, por seis profesores.

4.2. La prueba de Juan Pedro

En la prueba, constaban algunos datos personales del alumno, como edad, relación familiar y desempeño escolar (11 años, huérfano y precisa obtener nota mínima 5,0 para ser aprobado), según el Anexo B.

De forma análoga a la prueba de María Clara, la prueba de Juan Pedro fue corregida por diez profesores, conforme criterios individuales y sin previo debate. Enseguida, después de establecer de forma conjunta los criterios, se realizó la segunda corrección de la prueba. El Cuadro 5 presenta una síntesis comparativa entre la primera y segunda corrección de la prueba de Juan Pedro.

Cuadro 5 – Resumen de las correcciones de la Prueba de Juan Pedro

Ítem	1ª Corrección	2ª Corrección	Variación	Comentarios
Nota Mínima	3,0	4,0	1,0	Nota mínima subió de 3,0 para 4,0.
Nota Máxima	9,0	8,65	-0,35	Nota máxima se redujo de 9,0 para 8,65.
Nota Promedio	5,73	6,51	0,78	El promedio subió de 5,73 para 6,51.
Variación entre Nota Mínima y Máxima	6,0	4,65	-1,35	La variación entre la nota mínima y máxima cayó de 6,0 para 4,65.
Desvío Patrón	1,96	1,39	-0,57	El desvío patrón (dispersión en torno del promedio) bajó de 1,96 para 1,39.

Fuente: Los autores (2014).

Como se puede observar, la segunda corrección fue más consistente que la primera. Del punto de vista matemático, queda claro por la variabilidad de los datos, que:

- 1) La variación entre la nota mínima y máxima (amplitud) se redujo en 1,35, pasando de 6,0 para 4,65; y
- 2) El desvío patrón, que mide la dispersión en torno del promedio, se redujo en 0,57, pasando de 1,96 para 1,39.

Además, como beneficios complementares que se obtuvieron con la homogeneización del proceso de corrección, se obtuvo una elevación del promedio de la nota en 0,78, pasando de 5,73 para 6,51. Otra información importante, es que si se considerara la nota 5,0 (cinco) como siendo la mínima para aprobación, el alumno Juan Pedro por la primera corrección realizada por los profesores, sería aprobado por sólo seis de ellos; por la segunda corrección, por ocho profesores.

4.3. Discusión de los resultados

La discusión de los criterios para la segunda corrección de la prueba y la búsqueda de consenso en los aspectos evaluados trajeron un ensanchamiento del proceso evaluativo y una estandarización esmerada y constructiva, que reduce la subjetividad.

Esto quedó claro cuando se identificaron las reducciones de las oscilaciones entre nota mínima y máxima en las correcciones de las mismas pruebas. En la prueba de la alumna María Clara, la reducción fue de 1,5 puntos, pues la variación en la primera corrección era de 5,5 puntos, pasando para 4,0 puntos en la segunda. De forma similar, en las correcciones de la prueba de Juan Pedro se obtuvo reducción de 1,35 puntos en la amplitud, pues pasó de 6,0 para 4,65.

La reducción del desvío patrón que tiene por consecuencia una mayor concentración en torno del promedio, fue otro indicativo de buenos resultados. En la prueba de María Clara la reducción fue de 0,27 (pasó de 2,01 para 1,74), y en la prueba de Juan Pedro, el resultado fue todavía mejor, la reducción fue de 0,57, pasando de 1,96 para 1,39.

Como resultados indirectos se obtuvo la elevación en los promedios. En la primera corrección de la prueba de María Clara el promedio era de 4,35 y en la segunda, 5,71 (se le agregó 1,36 punto). En el caso de la prueba de Juan Pedro el promedio aumentó en 0,78 décimos de punto de la primera para la segunda corrección, pues fueron, respectivamente, 5,73 y 6,51. Además, si se considera el valor 5,0 (cinco) como nota mínima para la aprobación, María Clara sería aprobada por sólo cuatro profesores en la primera corrección, y por seis en la segunda. Ya Juan Pedro en la primera corrección sería aprobado por seis, y en la segunda, por ocho profesores.

Otro punto importante fue “no considerar” los planteos personales (sexo, edad, condiciones socio-económicas, relación familiar y desempeño anterior) en la evaluación, lo cual estimuló una mayor imparcialidad y sentido de justicia; aunque no siempre esa virtud sea fácilmente realizable, pues la racionalidad no es absoluta en el ser humano, influenciado por las variaciones momentáneas de humor y emociones.

De difícil medida, los aspectos personales de los alumnos pueden haber impactado en el proceso de atribución de la nota de la prueba. Un comentario de un profesor fue “Le tuve pena a Juan Pedro, pues precisaba 5,0 para ser aprobado, por eso facilité la corrección para obtención de la nota mínima necesaria para su aprobación”.

Finalmente, la elaboración de planteos o preguntas debe ser esmerada y muy bien evaluada, evitando malentendidos y ofreciendo la oportunidad para que el estudiante muestre su nivel de aprendizaje. Camargo (2011) enfatiza que la planificación de una prueba es fundamental para que se la elabore bien. Las preguntas deben ser claras, adecuadas al nivel de los alumnos, coherentes con las clases dadas y bien contextualizadas.

5. Consideraciones Finales

La evaluación es un proceso inherente a la actividad educacional por medio del cual se obtiene informaciones que permiten la toma de decisiones en varios niveles. La evaluación en el aula tiene por objeto principal verificar el aprendizaje de los alumnos, a menudo, promoviéndolos para el nivel siguiente, o reteniéndolos en el mismo nivel, cuando no hayan obtenido el mínimo conocimiento para su avance. Además, por medio de la evaluación de los alumnos, se pueden identificar fallas en el aprendizaje, de modo puntual o genérico, dándole al docente la oportunidad de perfeccionar su estrategia de enseñanza para determinados temas, o incluso, como lo señala Luckesi (1998), permitiendo perfeccionar el proceso de enseñanza-aprendizaje.

De este modo, se hace necesario que la evaluación sea consistente con las estrategias de enseñanza-aprendizaje, debiendo ser compuesta por múltiples y diversos instrumentos de evaluación y no sólo por la tradicional prueba escrita. Cuánto más instrumentos de evaluación existan, cuánto más variados sean (trabajos de investigación, seminarios, paneles, debates, dramatización, discusiones, experiencias, diario de bordo,

dinámica de grupos, juegos, murales, visitas técnicas, prueba oral), mayor será el universo de informaciones del profesor para hacer su “juicio de valor” esmerado, imparcial y justo. Además, los múltiples instrumentos reducen la posibilidad de injusticias, sobre todo porque un alumno puede tener buen desempeño en un determinado tipo de actividad y en otro no. Suhr (2008) complementa la lista de técnicas e instrumentos para recoger datos sobre el aprendizaje, como: autoevaluación, evaluación recíproca de los alumnos, observación, argumentación, producción de textos, defensa de proyectos y trabajos en grupos.

En el caso de la prueba escrita, tan tradicional en nuestro país, que tiene su existencia perpetuada por el sistema documental que todavía fortalece en el sistema educacional brasileño, con apoyo incondicional de profesores, alumnos y padres de alumnos. La prueba escrita es en primer lugar un “documento”. Documento que demuestra (o cree demostrar) cuánto un alumno sabe sobre un determinado asunto: es registro materializado que sirve para efectuar o refutar contra argumentos!

La evaluación de un alumno en una determinada asignatura debería ser holística, continua, considerando también los instrumentos formales, la participación del alumno, el interés demostrado, sus lecturas, investigaciones y todos los trabajos desarrollados.

No se puede tener una misma prueba evaluada por un profesor con nota 2,0 y por otro, con nota 7,5 – como fue el caso de María Clara. Peor todavía, un profesor evaluar la prueba de Juan Pedro con nota 3,0, y otro con nota 9,0. El establecimiento de patrones de corrección y documentados reduce las variaciones, es decir, es condición necesaria para la realización de una evaluación más esmerada, pero no es condición suficiente.

Los resultados de la investigación señalan la necesidad de una urgente reflexión crítica de profesores y coordinadores educacionales en el proceso de evaluación y de sus instrumentos, teniendo como prioridad el binomio “¿qué medir?” y “¿cómo medir?”

Si se quiere continuar con las pruebas escritas, es necesario que los profesores sean orientados y capacitados para saber cómo elaborar pruebas para medir el conocimiento, cómo definir y divulgar los criterios de evaluación. Es necesario también que los profesores tengan tiempo para realizar las correcciones de pruebas y trabajos, sean remunerados por dichas actividades de corrección de pruebas, para poder realizar el trabajo con criterio y asertividad. Por fin, la evaluación en el aula por medio de pruebas

escritas precisa ser revisada en un contexto más amplio, incluyendo la formación y capacitación de profesores.

Referencias

- CAMARGO, P. de. Professor posto à prova. *Revista Educação*, San Pablo, v. 174, set. 2011. Disponible en: <<http://revistaeducacao.uol.com.br/textos/174/artigo236301-1.asp>>. Accedió el: 19 jul. 2015.
- DEMO, P. *Praticar ciência: metodologias do conhecimento científico*. San Pablo: Saraiva, 2012.
- FERNANDES, D. *Avaliar para aprender: fundamentos, práticas e políticas*. San Pablo: UNESP, 2009.
- GATTI, B. A.; BARRETO, E. S. de S. (Coord.). *Professores do Brasil: impasses e desafios*. Brasília, DF: UNESCO, 2009. Disponible en: <<http://unesdoc.unesco.org/images/0018/001846/184682por.pdf>>. Accedió el: 19 jul. 2015.
- GUBA, E. G.; LINCOLN, Y. S. *Fourth generation evaluation*. Newbury Park: Sage Publications, 1989.
- HADJI, C. *A avaliação, regras do jogo: das intenções aos instrumentos*. 4. ed. Porto: Porto, 1994.
- INEP. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. *Censo Escolar da Educação Básica 2014*. Brasília, DF: INEP, 2014a. Disponible en: <<http://portal.inep.gov.br/basica-censo>>. Accedió el: 22 mar. 2015.
- _____. *Censo da Educação Superior 2013*. Brasília, DF: INEP, 2014b. Disponible en: <http://download.inep.gov.br/educacao_superior/censo_superior/apresentacao/2014/coletiva_censo_superior_2013.pdf>. Accedió el: 22 mar. 2015.
- LUCKESI, C. C. *Avaliação da aprendizagem escolar: estudos e proposições*. 8. ed. San Pablo: Cortez, 1998.
- PENNA FIRME, T. Avaliação: tendências e tendenciosidade. *Ensaio*, Río de Janeiro, v. 1, n. 2, p. 105-115, en./mar. 1994.
- SOUZA, C. P. de. Descrição de uma trajetória na/da avaliação educacional. *Série Ideias*, San Pablo, n. 30, p. 161-174, 1998.
- SUHR, I. R. F. *Processo avaliativo no ensino superior*. Curitiba: IBPEX, 2008.

Inaccurate Evaluation in the Correction of Written Exams Resulting from Lack of Pre-Defined Criteria

Abstract

Field research conducted with ten teachers who are students in a *stricto sensu* postgraduate course in Education, which aims to highlight the importance of having well-defined criteria in test evaluation. The research was based on a careful simulation of a test applied to two hypothetical elementary school students, which were graded by individual teachers. The first correction applied individual criteria; for the second correction, the group of teachers discussed and established the criteria and made another individual grading. The results showed significant improvements in the assessment, since the distortion of marks awarded was reduced. However, the results also indicate that the evaluation requires a critical reflection of teachers and educational coordinators, not only for test grading, but in the process of evaluation and its instruments, development issues, highlighting the importance of the binomial "what to measure?" and "how to measure?".

Keywords: Review. Test grading. Grading criteria.

A Imprecisão Avaliativa na Correção de Provas Escritas Decorrente da Falta de Critérios Pré-Definidos

Resumo

Pesquisa de campo realizada com dez professores que são alunos em um curso de pós-graduação *stricto sensu* em Educação, que visa evidenciar a relevância de se ter critérios bem definidos na avaliação de provas. A pesquisa constitui-se na simulação de uma prova realizada por dois alunos hipotéticos do ensino fundamental, que foi corrigida individualmente pelos professores. A primeira correção foi efetuada com critérios individuais; para a segunda correção, o grupo de professores discutiu, estabeleceu os critérios e efetuou nova correção individual. Os resultados apontam que a padronização de critérios reduziu as distorções de notas atribuídas. Todavia, os resultados ainda indicam que a avaliação exige uma reflexão crítica de professores e coordenadores educacionais, não somente quanto à correção de

provas, e sim, no processo de avaliação e de seus instrumentos, elaboração de questões, evidenciando a importância do binômio “o que medir?” e “como medir?”.

Palavras-chave: Avaliação. Correção de provas. Critérios de correção.

ANEXOS

Anexo A – Prova de Maria Clara

PROVA FINAL – 4ª. série do Ensino Fundamental

Nome: Maria Clara

Idade: 10 anos

Sexo: feminino de classe média.

Particularidades: _____

Cada questão vale 2,5 pontos.

Questão 1 – Utilizando as palavras: viagem, capital, hotel, passeio e cultura, escreva sobre uma viagem feita ao Estado de Santa Catarina.

A viagem com meus pais pela praia de Florianópolis foi legal, pois fizemos passeios e o hotel era legal.

Questão 2 – Hoje é aniversário de Daniela e Sérgio. Ele fez 13 anos e ela fez o dobro da idade dele. Quantas velinhas devo comprar para fazer a festa para eles.

4 velinhas velas velhinhas.

Questão 3 – Dom Pedro II foi imperador do Brasil por quase 50 anos no século XIX. Pergunta-se: qual cidade ele fundou no Estado do Rio de Janeiro e quando se deu o final do Império no Brasil?

Teresopolis. Terminou com a chegada da república

Questão 4 – Assinale no mapa abaixo, onde está a capital do país e o Rio Amazonas.

 → não fica muito perto não

NOTA FINAL : _____

Anexo B – Prova de João Pedro

PROVA FINAL – 4ª. série do Ensino Fundamental

Nome: João Pedro
Idade: 11 anos
Sexo: masculino
Particularidades: olhos e precisa de nota 5,0 para passar.

Cada questão vale 2,5 pontos.

Questão 1 – Utilizando as palavras: viagem, capital, hotel, passeio e cultura, escreva sobre uma viagem feita ao Estado de Santa Catarina.

Fui viajar com minha família para Blumenau que tem cultura alemã e foi um passeio muito ~~gosto~~ gostoso e legal.

Questão 2 – Hoje é aniversário de Daniela e Sérgio. Ele fez 13 anos e ela fez o dobro da idade dele. Quantas velinhas devo comprar para fazer a festa para eles.

Duas sendo uma para cada um.

Questão 3 – Dom Pedro II foi imperador do Brasil por quase 50 anos no século XIX. Pergunta-se: qual cidade ele fundou no Estado do Rio de Janeiro e quando se deu o final do Império no Brasil?

A cidade foi Petrópolis e o império terminou no final do século XIX.

Questão 4 – Assinale no mapa abaixo, onde está a capital do país e o Rio Amazonas.



NOTA FINAL : _____