

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Educación

**ESTUDIO DE LA VALIDEZ CONCURRENTES DE LOS INSTRUMENTOS
ELABORADOS POR LAS MAESTRAS DE UNA INSTITUCIÓN PRIVADA
MIXTA DE LA CIUDAD DE GUATEMALA
PARA EVALUAR EL APRESTAMIENTO A LA LECTURA**

ALVARO MAURICIO FORTIN MORALES



Guatemala
1998

**ESTUDIO DE LA VALIDEZ CONCURRENTES DE LOS INSTRUMENTOS
ELABORADOS POR LAS MAESTRAS DE UNA INSTITUCIÓN PRIVADA
MIXTA DE LA CIUDAD DE GUATEMALA
PARA EVALUAR EL APRESTAMIENTO A LA LECTURA**

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Educación

**ESTUDIO DE LA VALIDEZ CONCURRENTES DE LOS INSTRUMENTOS
ELABORADOS POR LAS MAESTRAS DE UNA INSTITUCIÓN PRIVADA
MIXTA DE LA CIUDAD DE GUATEMALA
PARA EVALUAR EL APRESTAMIENTO A LA LECTURA**


ALVARO MAURICIO FORTIN MORALES

**Trabajo de graduación presentado para optar al grado académico de Maestría en Medición,
Evaluación e Investigación en Educación.**

[Handwritten signature and scribbles]

**Guatemala
1998**

Vo. Bo.:

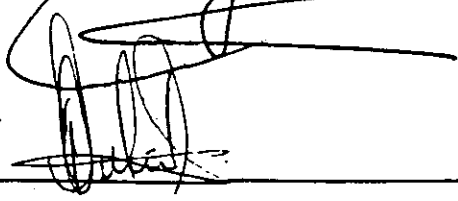
(f) 

Licenciado Fernando Rubio, MA
Asesor

Tribunal:

(f) *Harriet de Souza*

(f) *M. Moyano*

(f) 

Fecha de aprobación:

26 de septiembre de 1998.

CONTENIDO

		Pg.
I.	INTRODUCCIÓN.....	1
II.	PROBLEMA.....	2
III.	JUSTIFICACIÓN.....	3
IV.	OBJETIVOS.....	4
V.	MARCO CONTEXTUAL.....	5
VII.	FUNDAMENTACIÓN TEÓRICA	
	A. Evaluación y aprendizaje de la lectura.....	6
	B. Validez e instrumento criterio.....	13
VIII.	METODOLOGÍA	
	A. Participantes.....	18
	B. Diseño y método	18
	C. Instrumentos validados.....	20
	D. Instrumentos criterio.....	21
IX.	RESULTADOS.....	22
X.	DISCUSIÓN DE RESULTADOS.....	26
XI.	RECOMENDACIONES.....	30
XII.	IMPLICACIONES DEL ESTUDIO.....	34
XIII.	BIBLIOGRAFÍA.....	36
	APÉNDICES	
	A. Conceptos y terminología estadística.....	38
	B. Análisis de reactivos determinados en base a la muestra de Kindergarten de la institución del MRT-español Nivel 1.....	41
	C. Análisis de reactivos determinados en base a la muestra de Preparatoria de la institución del MRT-español Nivel 2.....	43
	D. Análisis de reactivos del instrumento elaborado por las maestras para medio año de Kindergarten.....	45
	E. Análisis de reactivos del instrumento elaborado por las maestras para fin de año de Kindergarten.....	47

LISTA DE TABLAS Y GRÁFICAS

Tabla #	Pg.
1. Contenido de la prueba MRT Nivel 1.....	16
2. Contenido de la prueba MRT Nivel 2.....	17
3. Coeficientes de confiabilidad (KR20) y errores estándar de medición de acuerdo al “Technical Supplement” del MER- español Nivel 1.....	17
4. Coeficientes de confiabilidad (KR20) y errores estándar de medición de acuerdo al “Technical Supplement del MRT-español Nivel 2.....	17
5. Contenido de la prueba de medio año de Kindergarten diseñada por las maestras.....	20
6. Contenido de la prueba de fin de año de Kindergarten diseñada por las maestras.....	21
7. Estadísticas descriptivas determinadas en base a la muestra de Kindergarten de la institución del MRT-español Nivel 1.....	22
8. Estadísticas descriptivas determinados en base a la muestra de Kindergarten de la institución del MRT-español Nivel 2.....	22
9. Estadísticas descriptivas del instrumento elaborado por las maestras para medio año de Kindergarten.....	23
10. Estadísticas descriptivas del instrumento elaborado por las maestras para fin de año de Kindergarten.....	23

RESUMEN

El trabajo que se presenta a continuación analiza la validez de los instrumentos elaborados por maestras de una institución privada mixta de la ciudad capital de Guatemala para evaluar el aprestamiento a la lectura de sus alumnos. La población estudiada para determinar la validez de los instrumentos mencionados está constituida por niños que van a iniciar la segunda parte de Kindergarten, y por niños que van a iniciar la Preparatoria. Estas pruebas se aplican de manera rutinaria tanto durante el mes de junio como durante el mes de octubre para evaluar el funcionamiento de los alumnos, y así determinar si éstos se beneficiarían de una promoción, de una retención o del ingreso a programas especiales de destrezas. El estudio es importante pues permite un mejor aprovechamiento de los recursos existentes.

Se discuten los contenidos de las evaluaciones a la luz de la investigación actual. Con base en ello se intenta determinar si las evaluaciones de las maestras consideran áreas significativas en relación al aprendizaje de la lectura. Se realiza también una comparación de los instrumentos desarrollados por las maestras de la institución, con instrumentos criterio, en este caso el MRT español en sus niveles 1 y 2.

Finalmente se discuten los hallazgos respecto a la validez de los instrumentos desarrollados por las maestras de la institución. Se hacen sugerencias para mejorar el instrumento y se analiza cuáles elementos debieran ser añadidos o eliminados. Así mismo, se hacen consideraciones respecto de la información recopilada y su implicación en la evaluación de otras poblaciones guatemaltecas.

I. INTRODUCCION

El presente estudio investiga la validez de los instrumentos realizados por maestras de una institución privada mixta de la ciudad capital de Guatemala (de aquí en adelante denominada solamente como "institución) para evaluar el aprestamiento a la lectura de alumnos que van a iniciar la segunda parte de Kindergarten, y de alumnos que van a iniciar la Preparatoria. El estudio es importante pues se requiere exactitud en las evaluaciones realizadas para aprovechar de manera óptima el tiempo y recursos disponibles. Para cumplir con el objetivo antes mencionado, se desarrollaron dos objetivos más específicos. El primero es determinar si los contenidos de las evaluaciones de las maestras se ajustan a lo que la teoría acerca del aprendizaje de la lectura propone. El segundo es comparar las evaluaciones de las maestras con un criterio ya establecido, en este caso las pruebas MRT.

Tras haber realizado el trabajo de aplicación de los instrumentos, el análisis produjo cuatro grupos de resultados. Dos de estos se refieren a la verificación de los MRT como instrumentos criterio para la población seleccionada y el segundo grupo compara las pruebas elaboradas por las maestras con los instrumentos criterio, o sea los MRT nivel 1 y 2. Además, se compara el contenido de las pruebas elaboradas por las maestras con el contenido que se ha determinado teóricamente y a través de investigaciones valiosas en la evaluación del potencial del aprendizaje de la lectura.

La comparación teórica antes mencionada sirve como paso previo a la determinación de la validez concurrente de los instrumentos elaborados por las maestras. El primer paso fue determinar la utilidad de los MRT. Se aplicaron estas pruebas a las mismas poblaciones a las que más tarde se les aplicó el instrumento sujeto a validación. Las puntuaciones obtenidas por los alumnos sirvieron para determinar si los instrumentos funcionan de manera adecuada para la población. Comprobado esto, sirvieron de criterio para comparación con las evaluaciones desarrolladas por las maestras.

II. PROBLEMA

El presente estudio investiga la validez de los instrumentos desarrollados por las maestras de la institución ya descrita en la sección previa para evaluar el aprestamiento a la lectura de sus alumnos. La población estudiada para determinar la validez de los instrumentos mencionados está constituida por niños que iniciarán la segunda parte de Kindergarten, y por niños que iniciarán la Preparatoria. Estas pruebas se aplican de manera rutinaria tanto durante el mes de junio como durante el mes de octubre para evaluar el funcionamiento de los alumnos, y así determinar si éstos se beneficiarían de una promoción, de una retención o del ingreso a programas especiales de destrezas.

III. JUSTIFICACION

El estudio es importante para la institución, la cual utiliza un programa que intenta responder a un doble grupo de exigencias (programas de EUA y Guatemala). Por ello, se requiere de exactitud en las evaluaciones realizadas para aprovechar de manera óptima el tiempo disponible para completar ambos programas. Evaluar para obtener datos que permitan realizar juicios con exactitud sobre el funcionamiento potencial de los niños permite a) adecuar los programas para proveer de soporte a aquellos niños que presentan dificultades; b) determinar con mayor propiedad cuándo son adecuadas las sugerencias de retención o promoción; c) verificar el desarrollo académico de los grupos involucrados y d) planear modificaciones o intervenciones del programa que ejecutan. Para satisfacer estas necesidades las maestras de la institución desarrollan pruebas que evalúan el rendimiento, de tal manera que con la mayor exactitud posible logren determinar si han cumplido con los criterios que se propusieron llenar en sus objetivos. Sin embargo, aún no se ha realizado un estudio acerca de la validez de tales evaluaciones para verificar que éstas llenen los requisitos necesarios para concluir que son herramientas útiles en base a las cuales se pueden elaborar juicios justos.

IV. OBJETIVOS

Debido a que los instrumentos son desarrollados por las maestras, tomando en cuenta el programa que ellas mismas diseñan, es importante considerar si las pruebas en realidad cumplen el propósito de proveer información adecuada para realizar juicios acertados. Para cumplir con el **objetivo** antes mencionado, se desarrollaron dos objetivos más específicos. El primero es determinar si los contenidos de las evaluaciones de las maestras se ajustan a lo que la teoría acerca del aprendizaje de la lectura propone (validez de contenido). El segundo es comparar las evaluaciones de las maestras con un criterio ya establecido, en este caso las pruebas Metropolitan Readiness Tests (de aquí en adelante MRT).

V. MARCO CONTEXTUAL

El presente estudio analiza la validez concurrente de los instrumentos elaborados por las maestras de la institución para evaluar el aprestamiento a la lectura. La institución educativa en donde se desarrolló el estudio, abarca los grados comprendidos entre Prekindergarten y 5° Bachillerato en Ciencias y Letras. Su programa es bilingüe, bicultural y biprogramático. Es decir, los cursos son impartidos en inglés y español, e introducen al alumno tanto en la cultura guatemalteca como en la estadounidense y cubren tanto el programa aceptado por la “Southern Association of Schools and Colleges” de EUA como por el Ministerio de Educación de Guatemala. Su filosofía hace énfasis en educar alumnos bilingües con valores positivos, que puedan desempeñarse tanto dentro de la cultura estadounidense como dentro de la cultura guatemalteca.

La institución realiza evaluaciones anuales para determinar si los alumnos han llenado los requisitos establecidos por sus programas para decidir si deben ser promovidos o no al próximo grado, o si se beneficiarían de ingresar a programas especiales de ejercitación de destrezas. Estas evaluaciones toman distintos modelos a través del año de acuerdo a la preferencia de los maestros. Sin embargo, existen dos evaluaciones mayores, una realizada al terminar el primer semestre y otra al finalizar el año, que se aplican de manera grupal con formato de papel y lápiz.

El presente estudio concierne las evaluaciones de medio y de fin de año de los alumnos de Kindergarten, quienes son sometidos a esta evaluación para determinar su aprestamiento a la lectura. Si son promovidos entran a Preparatoria, donde se les somete de lleno al aprendizaje y ejercitación en las destrezas de lectura.

VI. FUNDAMENTACIÓN TEÓRICA

A. Evaluación y aprendizaje de la lectura.

Para cumplir con los propósitos ya enumerados en la sección de objetivos, es decir a) determinar si los contenidos de las evaluaciones de las maestras cumplen con los requisitos de la validez de contenido y b) comparar las evaluaciones de las maestras con un criterio ya establecido, es necesario considerar cuáles aspectos deben tomarse en cuenta para desarrollar un instrumento que evalúe cuán preparados están los alumnos para aprender a leer. Es importante considerar que la instrucción de la lectura no puede concebirse como un proceso estándar, del cual puede determinarse una única forma para su óptima enseñanza. Dos métodos que actualmente sobresalen por el interés que han despertado entre investigadores son las perspectivas de “proceso” y la “tradicional”. La perspectiva que acentúa el desarrollo de “procesos” es usualmente vinculada a la corriente del “lenguaje integrado”, aunque también existe un tipo más puro que acentúa estrategias (Stahl, Pagnucco & Suttles, 1996; Gunning, 1992). Asimismo, respecto de la perspectiva “tradicional”, existe cierta variabilidad. En general, los programas “tradicionales” contemplan la instrucción de manera organizada y ordenada, incrementando dificultad según un programa preestablecido, y usualmente de acuerdo con un libro de texto básico llamado también “texto basal” (Gunning, 1992) con énfasis en los aspectos fonéticos de la interpretación de la palabra escrita (Stahl, Pagnucco & Suttles, 1996). Sin embargo, actualmente se han incorporado a esta tendencia las nuevas ideas de la conciencia fonológica, las cuales por sus características se acoplan mejor a los métodos tradicionales de instrucción (Vellutino, 1991; Byrne & Filding-Barnsley, 1993). Sin embargo, su énfasis en una destreza específica, alrededor de la cual gira un gran número de investigaciones, justifica tomarla en cuenta de manera individual.

La perspectiva de procesos postula que la instrucción de la lectura debe darse de manera natural, respondiendo al medio, necesidades e intereses del estudiante. El aprendizaje de la lectura depende y requiere del aprendizaje de la escritura. Por ello se acepta una etapa de “ortografía inventada” en la cual el niño construye palabras sobre la

base de su conocimiento de la lectura. Esta etapa es considerada apropiada y es fomentada pues se cree que es el primer paso en la construcción de formas menos concretas, más formales, de conceptualizar la palabra escrita. No se considera importante la velocidad de la lectura, y la capacidad de pronunciar con exactitud es solo un objetivo secundario a la comprensión y capacidad de transmisión de ideas y conceptos (Stahl, Pagnucco & Suttles, 1996; Gunning, 1992).

Por otra parte, la perspectiva de procesos se concentra en procesos metacognoscitivos; en otras palabras, las formas más adecuadas de pensar sobre cómo se lee para autoevaluar el propio funcionamiento. El término metacognición se refiere al conocimiento y al control de las actividades del pensamiento y el aprendizaje (Woolfolk, 1987), “el conocimiento propio acerca de los propios procesos cognoscitivos y productos o cualquier aspecto relacionado con esto, por ejemplo, las propiedades relevantes del aprendizaje de la información” (trad.lib.) (Flavell, 1976). Se introducen estrategias específicas, como definir la idea principal y responder a tales preguntas como “cuándo, dónde, cómo” (Gunning, 1992) y “por qué” (Woolfolk, 1987). Se hacen estas referencias debido a que la metacognición comprende dos aspectos principales: 1) saber qué hacer, a qué se refiere la tarea, qué estrategias y habilidades se requieren para completarla y 2) tener la capacidad de utilizar mecanismos autorreguladores para completar la tarea, saber cómo y cuándo hacer las cosas (Woolfolk, 1987).

La utilización de estos mecanismos se conoce como monitoreo cognoscitivo e interviene en el flujo de información a través de los sistemas de memoria y procesamiento de información (Woolfolk, 1987). Al tratar acerca de la lectura, las ideas alrededor de la metacognición suponen que la comprensión de la lectura es el resultado de una interacción entre el texto y las estructuras cognoscitivas que el lector utiliza y aplica durante el proceso de lectura (Woolfolk, 1987). Así, no es suficiente el proceso de decodificar, el cual es una etapa inmadura. Más bien, se requiere que el lector intente abstraer un significado de la lectura. Es debido a la atención a procesos metacognoscitivos y al cuidado necesario en la autoevaluación, que esta forma de

aprendizaje de la lectura con frecuencia se ve vinculada con programas de “lenguaje integrado”.

Las tendencias descritas anteriormente tienen en común que implican un aprendizaje que supone el funcionamiento de destrezas cognoscitivas complejas y un alto nivel de motivación hacia el aprendizaje. Hasta el momento no se ha tratado la cuestión acerca de las destrezas más básicas. Es obvio que el aprendizaje a través del sistema ya descrito tiene ventajas. Se requiere que el alumno desarrolle estrategias de pensamiento que lo lleven al autodescubrimiento a través de un camino marcado por errores aceptables, los cuales en última instancia le conducen a modificar su conocimiento. Pero a pesar del apoyo proporcionado a estas perspectivas, las investigaciones que comparan el aprendizaje de la lectura a través de distintos métodos muestran una tendencia más favorable hacia los sistemas más tradicionales (Vellutino, 1991).

Sin embargo, es necesario hacer la observación que en este momento se hace referencia solamente al aprendizaje inicial de la lectura y no al incremento de habilidad en ésta. Tampoco se hace referencia a otras materias de estudio. Es probable que las ventajas de los sistemas más tradicionales en oposición a los inclinados por la ejercitación de funciones cognoscitivas disminuyan al incrementar el nivel académico. De hecho, las ventajas observadas en el desarrollo de la lectura respecto de métodos tradicionales, se advierten cuando se toma la conciencia fonológica como una variable relevante del sistema tradicional utilizado. Más aún, los estudios parecen mostrar que los métodos tradicionales funcionan mejor con alumnos principiantes en lectura, mientras las ventajas de las perspectivas de procesos y estrategias incrementan según el alumno adquiere maestría en las destrezas básicas (Vellutino, 1991).

Esto puede responder al nivel de lectura de los alumnos. Para los niños que inician el aprendizaje de la lectura, el aprendizaje de los símbolos y su asociación a sonidos o a conceptos no se da de manera espontánea. El logro de la destreza de lectura alfabética requiere que se desarrollen representaciones mentales de los fonemas y de las asociaciones de cómo las letras representan a los fonemas. Esto se logra de manera más eficiente a través de actividades específicamente diseñadas para esto (Ehri, 1993). Es

sobre la base de estas ideas que la forma de evaluación respecto al aprendizaje, logro o potencial de la lectura diferirá. De tal manera que es posible concluir que si hasta el momento se ha encontrado evidencia de que el aprendizaje inicial de la lectura depende del aprendizaje de elementos fonéticos y de la representación del lenguaje, la evaluación deberá responder a estas necesidades.

Como se mencionó al inicio de este capítulo, la perspectiva “tradicional” propone una instrucción organizada según un programa preestablecido. Dentro de esta tendencia se encuentran tales programas como el de la “palabra generadora”, el método global, el silábico, el onomatopéyico y otros. La discusión de cada uno de estos correspondería a un texto general de técnicas de instrucción de lectura. Para propósitos de este trabajo basta indicar que los resultados que han producido, y la prevalencia y duración en las escuelas de tales programas han sido de gran variabilidad (Gunning, 1992; Stah, Pagnucco & Suttles, 1996).

Recientemente la investigación respecto de una de las tendencias “tradicionales” ha tendido a mostrar resultados más constantes tanto a través del tiempo como de diversas poblaciones, esta es la conciencia fonológica (Vellutino, 1991; Byre & Filding-Barsley, 1993). Esta se diferencia de otras propuestas “tradicionales” por su énfasis en una destreza específica (la conciencia fonológica), la cual puede subdividirse en subdestrezas que serán analizadas en posteriores párrafos de esta misma sección. La teoría ha mostrado ser de gran utilidad pues no sólo proporciona directrices de la instrucción de la lectura, sino también provee de herramientas para evaluar cuán preparado está un individuo para iniciar el aprendizaje de la lectura. Asimismo, ayuda a identificar qué áreas requieren refuerzo.

Al considerar en qué consiste la “conciencia fonológica”, en comparación a otras perspectivas “tradicionales”, dos elementos pueden mencionarse que resultan relevantes para el presente estudio. En general, se acepta que el alumno debe ser capaz de asociar un sonido con una representación gráfica. Pero mientras otras perspectivas han considerado que la representación gráfica es indispensable para la asociación fonética, por lo que se inicia la instrucción con dibujos y representaciones variadas, la conciencia

fonológica asume que el individuo es capaz de distinguir la existencia de los diferentes fonemas o sonidos independientemente de su representación. Aún más, considera que es posible manipular los distintos sonidos que constituyen una palabra aún antes de saber escribirlos (Vellutino, 1991; Gough & Larson, 1997). Es la habilidad en estos dos aspectos lo que determina la capacidad del alumno para adquirir la habilidad de leer.

En función del presente trabajo, la evaluación de los instrumentos realizados por las maestras estará orientado hacia los criterios descritos en el párrafo anterior como característicos de la conciencia fonológica. Aún así, debe tomarse la precaución de no suponer que estos elementos son únicos y suficientes para el desarrollo adecuado de la lectura. Más bien, son considerados como elementos indispensables y básicos, pero no suficientes para el logro de la obtención de maestría. Es decir, es sólo un aspecto, el cual es considerado básico, pero sobre el cual se deberá construir un edificio más amplio de destrezas para el logro de la maestría en lectura.

Hasta este momento se han discutido algunas ideas con respecto de los métodos de enseñanza de la lectura. Tener información de esto es importante para planificar la evaluación, pues es sobre la base de lo que se haya introducido en clase que ésta debería plantearse. El sistema que escojamos tendrá impacto en nuestros objetivos, en la selección del material a cubrir, en los aspectos que consideramos importante evaluar. La evaluación en sí es un tema tan complejo como el método para enseñar. Según algunos, la evaluación de la lectura involucra un monitoreo constante concentrado en descripciones cualitativas, no necesariamente estandarizadas, prestando atención a debilidades y fortalezas específicas mostradas por los niños (Gunning, 1992; Van Kraayenoord, 1993). Sin embargo, junto a sus obvias ventajas (posibilidad de llevar un seguimiento más cercano, evitar que se establezcan hábitos que deterioren el funcionamiento, posibilidad de intervenir cuando los problemas o dificultades en la lectura recién se presentan), este sistema también presenta ciertas limitaciones. Esto es cierto en particular con respecto al tiempo y recursos necesarios para implementar el sistema, tanto a nivel puramente material como respecto de la preparación de las maestras involucradas en la evaluación.

En suma, el punto de vista ya presentado prescribe que la evaluación de la lectura, así como de cualquier otro tipo de aprendizaje, debiera ser constante y tan individualizada como sea posible cuando esto resulta accesible. Es gracias a estas mediciones que es posible tomar decisiones y realizar juicios mejor informados respecto al alumno. Esto, por supuesto, implica una situación ideal que no es realista para muchos individuos en la mayoría de ocasiones. De hecho, es prácticamente imposible cuando el objetivo es realizar estudios que permitan comparar poblaciones o determinar criterios mínimos de rendimiento. Por ello la evaluación estandarizada resulta más eficiente cuando se requiere conocer, con mínima inversión de recursos, el funcionamiento de un individuo específico en un momento determinado.

Como ya se señaló antes, los aspectos propuestos por la perspectiva tradicional y de conciencia fonológica parecen explicar en mayor medida el desempeño del niño en la lectura en su etapa inicial. Es por ello que este estudio trabaja sobre la base de la evaluación de esos aspectos como criterios del funcionamiento del aprendizaje **inicial** en lectura. Esto por sí mismo no determina qué factores específicos van a ser considerados. Existen diferentes puntos de vista respecto de qué variables son adecuados predictores en lectura. Respecto de lo anterior, es necesario tomar en cuenta aquellos factores que demuestran tener efecto sobre el aprendizaje de la lectura de acuerdo a investigaciones o estudios previos. Brian Byrne y Ruth Fielding-Barnsley (1993) realizaron un estudio que intentó evaluar el funcionamiento de niños sujetos a programas de lectura siguiendo el modelo de la conciencia fonológica. El grupo experimental recibió instrucción directa en identificar fonemas, mientras que el grupo control no. Los resultados mostraron que los niños que poseían un alto nivel de ejecución en la identificación de fonemas obtenían mejores puntuaciones en tareas relacionadas con reconocimiento de palabras incompletas, conocimiento del alfabeto, identificación de palabras, ortografía y reconocimiento de pseudopalabras. Además, el conocimiento alfabético predecía mejor el futuro desarrollo literario, pero la conciencia fonológica justificaba mejor la varianza en las mediciones de ortografía y decodificación. Por otra parte, la inteligencia verbal no parecía influir de manera sobresaliente sobre la ejecución en la lectura. Es importante, sin embargo, hacer

notar que estos resultados, así como la mayor parte de investigación desarrollada respecto de la instrucción de la lectura, proviene del aprendizaje de la lectura en idiomas distintos al español (como el inglés).

En el caso específico del español los resultados podrían variar, dado que la escritura de la lengua responde a un sistema fonético más simple. Sin embargo, la evidencia sugiere que el aprendizaje de la lectura en cualquier lengua cuyo sistema de escritura corresponde a un sistema ortográfico (español, inglés, portugués, etc.) responde de manera más adecuada a la instrucción en conciencia fonológica, en contraposición a uno ideográfico, en el cual no hay congruencia o relación entre los grafemas y los fonemas que construyen el lenguaje oral (por ejemplo, el chino) (Gough, Larson & Yopp, 1977).

Tal parece que aún en sistemas ortográficos simples, un solo fonema no siempre responde a un solo carácter (Lightbown & Spada, 1993). Este interactúa con factores puramente ortográficos, los cuales ganan peso al incrementar el nivel de dificultad del material que el lector utiliza. En suma, para lectores con un nivel inicial aún, el funcionamiento en conciencia fonológica parece explicar mejor el rendimiento actual y futuro (Bjälid & Lundberg, 1995).

Así pues, la evaluación de la lectura siguiendo estos lineamientos requeriría conocer cuán bien pueden los alumnos realizar tales identificaciones y asociaciones entre los grafemas y los fonemas. A un nivel de prelectura resultaría imposible esperar que tales asociaciones de símbolos ya estén establecidos firmemente. Más bien, es de esperar obtenerlos a través de símbolos alternos convencionalmente adquiridos y que puedan reflejar de alguna manera tal funcionamiento. Así mismo, se espera que el niño pueda realizar tareas de conciencia fonológica, tales como reconocer los sonidos que forman distintas palabras, rimar, reconocer sonidos iniciales de una palabra y reconocer sonidos finales de una palabra (Byrne & Fielding-Barnsley, 1993). Gough, Larson & Yopp (1977) enumeran varias tareas que pueden utilizarse para evaluar las destrezas relacionadas con la conciencia fonológica. Estas pueden clasificarse en: discriminación auditiva, la cual consiste en distinguir qué fonema se pronuncia; incorporación fonética,

es decir poder agregar un nuevo fonema a un vocablo que se le proporciona al niño de manera oral; conteo fonético, poder determinar con exactitud con cuántos fonemas cuenta una palabra enunciada oralmente; eliminación fonética, repetir una palabra que se le ha dado al estudiante eliminando alguno de los fonemas según instrucción previa; y segmentación fonética, o sea poder repetir la palabra fonema por fonema sin mezclarlos o combinarlos. Todas éstas se refieren a la capacidad que muestra el individuo para manipular los segmentos estructurales de la lengua hablada.

B. Validez e instrumento criterio.

Además de lo discutido hasta este momento, es necesario considerar otros asuntos relevantes del estudio. En este caso se intenta establecer la “validez” del instrumento diseñado por las maestras de la institución. Ese instrumento intenta predecir el funcionamiento del niño durante la fase escolar que corresponde al período durante el cual se introduce el aprendizaje de la lectura. En este sentido, “validez” puede definirse como “el término...que denota la utilidad científica de un instrumento de medida, en el que puede establecerse ampliamente qué tan bien mide lo que pretende medir” (Nunnally & Bernstein, 1995). Es importante notar que no es la prueba en sí la que es válida o inválida, sino más bien el uso que de ella se hace (Capper, 1996). Es por ello que la determinación de la validez de los instrumentos responde también a los propósitos planteados para los instrumentos. En otras palabras, utilizar un instrumento para evaluar lectura para juzgar si un estudiante es bueno en matemáticas es inapropiado e invalida la prueba, a pesar de que ésta sea muy útil para determinar el logro en lectura.

En general el término “validez” alude a tres significados principales: 1) validez de constructo, 2) validez de criterio: concurrente y predictiva y 3) validez de contenido (Nunnally & Bernstein, 1995; Remmers, Gage & Francis-Rummel, 1965; Crocker & Algina, 1986; Gronlund, 1993). La validez de constructo utiliza el término constructo como una variable abstracta y latente estimada o “construida” para reflejar una hipótesis, y la cual no posee una dimensión observable concreta y directa (Nunnally & Bernstein,

1995). La validez de contenido se refiere a cuán bien el contenido del instrumento representa el área o materia evaluada (Capper, 1996). La validez concurrente se evalúa mostrando hasta qué grado las puntuaciones de la prueba corresponden a una medida ya aceptada de ejecución o “estatus” que se realizó al mismo tiempo (Remmers, Gage & Francis Rummel, 1965).

Los tres tipos de validez pueden estar presentes para un mismo instrumento en diferentes grados, aunque idealmente debe haber una buena dosis de cada una de ellas. Sin embargo, puede haber un énfasis en alguno de los tipos específicos debido al punto en el cual se encuentra el instrumento en su desarrollo, o debido a los recursos existentes para su estudio. Los estudios de validez de constructo toman lugar usualmente cuando no existe un criterio o universo completamente aceptable, por lo que el diseñador de la prueba desea realizar una inferencia a partir de las puntuaciones agrupándolas bajo alguna etiqueta particular de algún constructo psicológico; la validez de criterio se utiliza en situaciones en las que el diseñador de la prueba desea realizar inferencias con base en las puntuaciones de los evaluados y a alguna variable real de importancia; y la validez de contenido se utiliza cuando el evaluador desea inferir a partir de las puntuaciones del examinado a un grupo más amplio de esos mismos aspectos evaluados, o sea si los reactivos presentados son una muestra representativa del constructo evaluado (Crocker & Algina, 1986).

En párrafos anteriores se realiza una breve descripción de algunas de las ideas actuales acerca de los aspectos tratados por la evaluación de las maestras. Sin embargo, el verdadero propósito del estudio es determinar la validez concurrente de los instrumentos elaborados por las maestras y los MRT. Se realizaron comparaciones entre las mediciones realizadas por medio de los instrumentos de las maestras y de los MRT, sirviendo el MRT como un criterio concurrente para determinar la validez de las evaluaciones de las maestras. Este tipo de comparaciones ha sido utilizado previamente por otros autores. Por ejemplo, un estudio en el que se comparó el MRT con evaluaciones basadas en currículo mostró que el MRT proporcionaba resultados

compatibles, aunque la riqueza de información era mayor a través del instrumento basado en el currículo (Joyce & Wolking, 1987).

El análisis factorial del nivel 2 de la prueba MRT (Watkins & Wiebe, 1984) ha mostrado que existe un factor general, pero fue necesario realizar una rotación varimax para elucarlo con mayor claridad. Tras la rotación se encontraron dos factores, uno compuesto por las tareas visuales, auditivas y cuantitativas, y otro factor conformado por las tareas de lenguaje escolar y comprensión auditiva. No se encontraron evidencias de factores específicos para cada una de las subáreas de las pruebas. Contrario a lo propuesto por la prueba, la cual plantea que el MRT en sí (restando la sección cuantitativa para el nivel 2) puede ser utilizado como una medida de prelectura, el estudio antes mencionado no encontró una clara medición de prelectura, a pesar de que sí se encontró que la prueba puede utilizarse para determinar el nivel de aprestamiento para ingreso a aquellos niveles en donde se inicia el aprendizaje de la lectura. De hecho, como otros autores han observado, (Clancy & Pianta, 1993) la prueba parece predecir adecuadamente el funcionamiento futuro del alumno desde el momento en que la prueba fue aplicada hasta segundo grado. Sin embargo, estas relaciones no se limitaron a aspectos puramente académicos, sino a otros que predisponen una buena ejecución escolar, como aspectos socio-emocionales. Además, incluye mediciones cognoscitivas y de funcionamiento académico, las cuales mostraron relaciones positivas con el funcionamiento escolar.

Deben tomarse en cuenta también algunos otros factores. Como lo describe el suplemento técnico de los instrumentos, las pruebas son una traducción, la cual sin embargo fue sometida a análisis para asegurar la pertinencia de los reactivos para la nueva población. Nuevos estudios psicométricos fueron llevados a cabo para la traducción. Es importante notar que la traducción de la prueba consistió en realidad en traducción de las instrucciones de las tareas que el niño debía realizar, (Nurss & Gauvran, 1993). Los reactivos en sí están constituidos por instrucciones impartidas oralmente (las cuales fueron traducidas por la compañía editora del instrumento) con opciones constituidas por representaciones gráficas. Durante la prueba, cada reactivo es

contestado por los alumnos después de que la maestra ha dado una instrucción específica por reactivo. Sin embargo, también se realizó un análisis de los datos obtenidos por los estudiantes de Kindergarten y Preparatoria en los MRT en la misma población de la institución para determinar su utilidad como instrumentos criterio para esta población.

La evidencia tiende a mostrar que los MRT pueden ser utilizados como instrumentos criterio adecuados para el propósito de este trabajo (ver tablas 3 y 4). Los estudios de Nurss & McGauvran (1993) indican que pueden utilizarse por hispanoparlantes no estadounidenses. Además, aunque también evalúa aprestamiento académico general, las destrezas de prelectura son específicamente tratadas por el instrumento (Nurss & McGauvran en Clancy & Pianta, 1993).

En las tablas 1 y 2 pueden observarse las áreas evaluadas por el MRT, tanto en el nivel 1 como en el nivel 2. En las tablas 3 y 4 se observan los datos suministrados por el manual de "The Psychological Corporation".

Tabla 1: Contenido de la prueba MRT Nivel 1.

	<u>ÁREA EVALUADA</u>	<u>NÚMERO DE REACTIVOS</u>
A.	DESTREZAS AUDITIVAS	
1.	Memoria auditiva	12
2.	Sonidos iniciales	14
B.	DESTREZAS VISUALES	
3.	Reconocimiento de letras	11
4.	Semejanzas visuales	14
C.	DESTREZAS DEL LENGUAJE	
5.	Lenguaje escolar y comprensión auditiva	15
6.	Lenguaje cuantitativo	11
	TOTAL	66

Tabla 2: Contenido de la prueba MRT Nivel 2.

	<u>ÁREA EVALUADA</u>	<u>NÚMERO DE REACTIVOS</u>
A.	DESTREZAS AUDITIVAS	
1.	Silabas iniciales	13
2.	Correspondencia entre sonido y silaba	16
B.	DESTREZAS VISUALES	
3.	Semejanzas visuales	10
4.	Encontrar modelos	16
C.	DESTREZAS DEL LENGUAJE	
5.	Lenguaje escolar	9
6.	Comprensión auditiva	9
D.	DESTREZAS CUANTITATIVAS	
7.	Conceptos cuantitativos	9
8.	Operaciones cuantitativas	15
	TOTAL	78

Tabla 3: Coefficientes de confiabilidad (KR20) y errores estándar de medición de acuerdo al "Technical Supplement" del MRT-español Nivel 1.

Kindergarten	Nivel 1
<i>N</i>	181
<i>Media</i>	52.9
<i>SD</i>	13.1
<i>R</i>	.93
<i>SEm</i>	3.6

Tabla 4: Coefficientes de confiabilidad (KR20) y errores estándar de medición de acuerdo al "Technical Supplement" del MRT-español Nivel 2 por grado.

Kindergarten	Nivel 2
<i>N</i>	150
<i>Media</i>	42.9
<i>SD</i>	11.9
<i>R</i>	.90
<i>SEm</i>	3.8
Preparatoria	Nivel 2
<i>N</i>	430
<i>Media</i>	53.6
<i>SD</i>	10.8
<i>R</i>	.90
<i>SEm</i>	3.4

VII. METODOLOGÍA

A. Participantes.

La muestra estuvo compuesta por los alumnos de Kindergarten y Preparatoria de la institución. No se tomó en cuenta a los niños repitentes pues en esos casos faltaba alguna de las dos puntuaciones de las evaluaciones utilizadas. Otros casos también fueron descartados por no contar con una o ambas de las puntuaciones requeridas por el estudio. El grupo de Kindergarten estuvo integrado por 68 participantes de una promoción de 73 niños. El 60% del grupo estuvo constituido por niñas, mientras que el 40% por niños. La edad promedio en enero de 1997 fue de 5 años 4 meses para este grupo, con una desviación estándar de 4 meses. En el caso de la Preparatoria el grupo estuvo conformado por 70 participantes lográndose en este caso tomar en cuenta 65 estudiantes para participar en el estudio. La proporción de los sexos fue de 58% de niños y 42% de niñas. La edad promedio en enero de 1997 fue de 6 años 4 meses con una desviación estándar de 4.7 meses.

B. Diseño y método.

El presente estudio pretende determinar la validez concurrente de los instrumentos elaborados por las maestras de la institución. Se tomaron los resultados del MRT correspondiente al grado en cuestión y se compararon con los resultados en la prueba elaborada por las maestras para el mismo grado. Los MRT fueron administrados en enero de 1997, cuando los alumnos de Kindergarten y Preparatoria se encontraban en el primer semestre de clases. Los datos fueron ingresados en una hoja electrónica de Microsoft Excel. El análisis de reactivos se realizó con el programa Systat 7.0 y el resto de los análisis en SPSS 7.0.

Se eliminaron los reactivos que evalúan el conocimiento de la correspondencia entre sonido y sílaba para el MRT nivel 2, así como la sección de reconocimiento de letras del MRT nivel 1. La razón para ello fue que estos contenidos no habían sido

impartidos al momento en que los niños fueron evaluados. De hecho, en años anteriores se había realizado la evaluación completa obteniendo resultados que según reportes de las maestras eran significativamente inferiores al resto de áreas. Además, elevaban los niveles de frustración de la población lo cual provocaba conductas disruptivas por parte de los examinados. Lamentablemente no se cuenta con datos numéricos para estas observaciones. Sin embargo, se consideró que la incompatibilidad con respecto al programa, y el hecho de que los instrumentos sometidos a validación no contenían este tipo de actividades, eran suficiente justificación para eliminar la sección. Fue debido a esto que también se optó por realizar un análisis de reactivos de las pruebas MRT como una forma de asegurar la calidad de los contenidos evaluados.

También se realizó un análisis de los resultados obtenidos por los estudiantes de Kindergarten y Preparatoria en los MRT en la misma población de la institución para determinar su utilidad como criterios e intentar identificar si había habido algún efecto negativo por omitir las secciones de “reconocimiento de letras” y “correspondencia entre sonido y sílaba”. De demostrar un adecuado funcionamiento, el nivel 1 del instrumento sería utilizado para determinar la validez de la prueba de medio año desarrollada por las maestras, mientras que el nivel 2 se utilizaría para determinar la validez de la prueba de fin de año. Los datos proporcionados por “The Psychological Corporation” son presentados en las tablas 3 y 4. Los datos para la población de la institución son presentados en la sección de resultados en las tablas 7 y 8. Los datos específicos respecto al análisis de los reactivos puede revisarse en los apéndices B, C, D y E.

Las pruebas desarrollada por las maestras estaban planificadas para Kindergarten. Sin embargo, una de éstas fue desarrollada para aplicarse a fin de año y la otra a medio año. Para el presente estudio se tomaron las pruebas de medio año aplicadas en 1997 y las de fin de año de 1996. Las primeras fueron contrastadas con el MRT para Kindergarten (nivel 1) mientras que las segundas fueron contrastadas con el MRT para Preparatoria (nivel 2), tomando en cuenta los reactivos referentes a lectura. Además, también en este caso las pruebas fueron analizadas por reactivo.

C. Instrumentos validados.

Los instrumentos sujetos a validación fueron diseñados por las maestras de Kindergarten, sección español de la institución. Estos fueron basados en el contenido del curso, y en las destrezas que según la experiencia de las maestras requieren los niños al continuar su progreso dentro de la secuencia de grados de la institución.

Las pruebas fueron diseñadas para ser aplicadas en Kindergarten, una a medio año y la otra a fin de año. Ambas fueron construidas de tal manera que la administración de reactivos es oral. Las opciones para cada reactivo se encuentran dibujados en un folleto, en el cual el niño marca su selección de respuesta. La prueba de medio año de Kindergarten cuenta con 54 reactivos distribuidos a lo largo de tres grandes secciones. Estas son percepción visual, percepción auditiva y destrezas de pensamiento. El instrumento para fin de año de Kindergarten cuenta con 75 reactivos distribuidos en dos grandes áreas. Estas son prelectura y matemáticas. En las tablas 5 y 6 puede observarse la tabla de especificaciones para cada uno de los instrumentos.

Tabla 5: Contenido de la prueba de medio año Kindergarten diseñada por las maestras.

	<u>AREA EVALUADA</u>	<u>NUMERO DE REACTIVOS</u>
A.	PERCEPCIÓN VISUAL	
1.	Discriminación forma	3
2.	Discriminación de posición	3
3.	Discriminación por detalle	3
4.	Copia de patrones	3
5.	Completación de figuras	9
6.	Seguimiento de patrones	3
B.	PERCEPCIÓN AUDITIVA	
7.	Palabra larga-corta	4
8.	Seguimiento de instrucciones	6
9.	Memoria auditiva a corto plazo	4
10.	Rimas	4
C.	DESTREZAS DE PENSAMIENTO	
11.	Categorización	4
12.	Análisis	4
13.	Secuencia lógica	4
	TOTAL	54

Tabla 6: Contenido de la prueba de fin de año Kindergarten diseñada por las maestras.

	<u>AREA EVALUADA</u>	<u>NUMERO DE REACTIVOS</u>
A. PRE-LECTURA		
1.	Asociación sonido-símbolo	10
2.	Dictado de vocales	5
3.	Sonido inicial	10
4.	Copia de diseños	10
5.	Figura-fondo	5
B. MATEMÁTICA		
6.	Conteo	5
7.	Dictado de números	10
8.	Secuencia numérica (antes-después)	10
9.	Adición	5
10.	Patrones	5
	TOTAL	75

D. Instrumentos criterio.

Los instrumentos criterio fueron analizados con los datos proporcionados por las mismas poblaciones a quienes se les aplicaron los instrumentos desarrollados por las maestras. Se realizó un análisis general del instrumento y un análisis por reactivo de su contribución a la prueba, tanto para los instrumentos criterio como para los instrumentos a ser validados. Estos fueron analizados en función de su confiabilidad individual, su correlación con respecto al resto de reactivos, la media y desviación estándar por cada uno. En este caso, se esperaría correlaciones moderadas para los reactivos. Así mismo, mientras más cercano a 1 se encuentre la media del reactivo individual, menor su dificultad. Estos resultados pueden observarse en los apéndices B, C, D y E.

VIII. RESULTADOS

Tras el análisis de los datos se obtuvieron cuatro series de resultados. Dos de éstas se refieren al funcionamiento de los MRT como instrumentos criterio para la población de la institución específicamente. Por otra parte, se obtuvieron resultados para las pruebas desarrolladas por las maestras, lo cual constituye el segundo objetivo del presente estudio. El primero de los objetivos fue inicialmente tratado en el “Marco Conceptual”, aunque su análisis más detenido se encuentra en la sección “Discusión”. A continuación se describen los resultados correspondientes a los datos obtenidos tras el análisis estadístico de los MRT en la institución y de los instrumentos desarrollados por las maestras. Asimismo, se analiza la correlación de estos instrumentos. En el apéndice A aparece una breve explicación de la terminología y conceptos estadísticos utilizados en esta sección.

A continuación, en las tablas 7, 8, 9 y 10, se presentan los resultados obtenidos tanto para las pruebas MRT, como para las desarrolladas por las maestras de la institución:

Tabla 7: Estadísticas descriptivas determinadas en base a la muestra de Kindergarten de la institución del MRT-español Nivel 1.

Media	45.957
n. de casos	70
Desviación estándar	8.486
Error estándar	1.022
Puntuación Máxima	64
Puntuación Mínima	28
Coefficiente Alfa	0.850

Tabla 8: Estadísticas descriptivas determinadas en base a la muestra de Preparatoria de la institución del MRT-español Nivel 2.

Media	56.609
n. de casos	69
Desviación estándar	9.609
Error estándar	1.165
Puntuación Máxima	73
Puntuación Mínima	34
Coefficiente Alfa	.849

Tabla 9: Estadísticas descriptivas del instrumento elaborado por las maestras para medio año de Kindergarten.

Media	43.800
n. de casos	55
Desviación estándar	5.303
Error estándar	0.748
Puntuación Máxima	54
Puntuación Mínima	27
Coefficiente Alfa	.764

Tabla 10: Estadísticas descriptivas del instrumento elaborado por las maestras para fin de año de Kindergarten.

Media	69.431
n. de casos	65
Desviación estándar	4.765
Error estándar	0.596
Puntuación Máxima	75
Puntuación Mínima	50
Coefficiente Alfa	0.812

La media para el nivel 1 de la institución se encontró por debajo de la de la muestra de validación original, pero es importante recordar que la muestra de la institución no estuvo expuesta a once de los reactivos originales. La desviación estándar fue menor y hubo un menor error estándar. Por otra parte, el coeficiente de confiabilidad resulta más favorable en la muestra original. Los resultados para el nivel 2 no mostraron medias o varianzas muy diferentes entre poblaciones, aunque en este caso los estudiantes de la institución no fueron expuestos a dieciséis de los reactivos originales. Los resultados fueron similares a aquellos encontrados en el nivel 1 respecto al índice de confiabilidad y al error estándar. Por otra parte, debe notarse que el tamaño de ambas muestras es significativamente diferente. En general, se tomaron los MRT como un adecuado instrumento para ser utilizado como criterio en este estudio.

Con respecto al análisis de los reactivos, resulta más adecuado considerarlos para cada prueba por separado. En el caso de la evaluación realizada para fin de año, la cual

fue contrastada con el MRT de Preparatoria, se obtuvo un coeficiente alfa de .812. La puntuación media fue de 69 puntos sobre 75, lo cual supera el rango de azar en las respuestas, pero presenta un techo más bajo que el del MRT. Ninguno de los reactivos produjo diferencias significativas en la confiabilidad de la prueba, tal como calculada por el coeficiente alfa, cuando era sustraído del resto de los reactivos. Sin embargo, de los 75 reactivos evaluados 42 de ellos mostraron un índice de confiabilidad cercano al 0 acompañados de bajas desviaciones estándar para el promedio del mismo reactivo dentro del grupo (media equivalente o muy cercana a 1.00 con desviaciones estándares menores de .17), lo cual parece indicar que no contribuyeron de manera significativa al valor discriminativo de la prueba.

En el caso de la prueba realizada para medio año, la cual fue contrastada con el MRT de Kindergarten, se obtuvo un coeficiente alfa de .764. La puntuación media fue de 43.80 puntos sobre 54, lo cual supera el rango de azar en las respuestas, pero presenta un techo más bajo que el del MRT, tal como el observado en el caso de la prueba de fin de año. En este caso la confiabilidad de la prueba se encuentra un tanto por debajo de aquella obtenida en el caso de la prueba de fin de año. Tampoco en este caso sin embargo, la ausencia de ninguno de los reactivos produjo diferencias significativas en la confiabilidad de la prueba calculada por el coeficiente alfa. De los reactivos evaluados 19 mostraron un índice de confiabilidad cercano al 0. Sin embargo en este caso solo 8 irían acompañados de bajas desviaciones estándar del promedio para el mismo reactivo dentro del grupo. Esto parece indicar que son estos 8 reactivos los que contribuyen en menor grado a la discriminación de los individuos altos y bajos. Por otra parte, existen otros 11 reactivos que poseen un bajo nivel de dificultad para ciertos individuos, pero no para otros. Por ello, parecieran constituir reactivos que podrían funcionar como criterio de discriminación.

En ambas pruebas se observa una tendencia de los reactivos desarrollados por las maestras a contar con niveles moderados o bajos de dificultad. En el caso del MRT hubo una mayor proporción de reactivos que fueron contestados correctamente por menos del 20% de participantes.

Respecto de los MRT puede observarse en los Apendices C y D que hubo un reactivo para el nivel 1 y 3 reactivos para el nivel 2 que se comportaron de manera contraria a lo esperado para un reactivo con adecuada capacidad de discriminación. Respecto de los reactivos del nivel 2, se observó que estos requerían solamente la discriminación del cambio de posición de dígitos en un caso, y de letras en el siguiente. Esto pudo haber provocado confusión para los alumnos del grupo alto, quienes en lugar de observar los estímulos como abstractos pudieron haber intentado descifrarlos. Por otra parte, para el nivel 1, puede observarse que los resultados se encontraron tan sólo ligeramente en dirección negativa. Es confuso en este caso decidir si realmente hubo un efecto contrario, o simplemente un efecto de 0 que por azar se desvió ligeramente hacia el sentido negativo de la correlación.

La correlación entre las puntuaciones de las maestras para la evaluación diseñada para el fin del ciclo de Kindergarten y el MRT correspondiente a Preparatorio produjo resultados significativos a un nivel alfa de .05 ($r=.303$; $p<.05$). Los resultados para la correlación entre la prueba de medio año para Kindergarten y el MRT correspondiente a Kindergarten produjo resultados significativos a un nivel alfa menor que .01 ($r=.329$; $p<.01$). En ambos casos el análisis fue de dos colas.

pruebas de las maestras contienen los elementos de reconocimiento de los elementos fonéticos en una palabra. Sin embargo, se evalúa poco la capacidad para manipular esos elementos.

La validez predictiva, concurrente o relacionada con el criterio se refiere al uso de un instrumento para estimar la conducta criterio que es externa al mismo instrumento de medición en el proceso de ser validado (Nunnally & Bernstein, 1995). En este caso es necesario tener alguna idea del funcionamiento de los niños en las tareas evaluadas por las pruebas desarrolladas por las maestras. Debido a razones prácticas, en este caso se optó por utilizar un instrumento externo, que ya hubiese sido sometido a análisis, para que funcione como tal.

Se seleccionó como instrumento criterio el MRT, en sus niveles 1 y 2 para utilizarse según el nivel apropiado. Se escogió este instrumento debido a que presenta una imagen global de las destrezas de los alumnos, procurando describir el nivel de adecuación actual para el rendimiento escolar. Esto es relevante pues el aprendizaje de la lectura es quizás el principal objetivo para la Preparatoria. La descripción de las áreas evaluadas por el instrumento pueden ser observadas en las tablas 1 y 2. Se debe notar que hubo ciertas áreas que no fueron tomadas en cuenta en esta oportunidad. Estas contenían elementos que las maestras no cubren en sus cursos. En años anteriores las maestras han observado funcionamiento muy pobre en esas áreas, acompañado de altos niveles de frustración por parte de los niños. Estos tendían a reflejarse en conductas disruptivas. Por estas razones se decidió omitir las secciones de "Reconocimiento de letras" para el nivel 1 y "Correspondencia entre sonido y sílaba" para el nivel 2. Se consideró apropiado por las razones mencionadas.

En este estudio el MRT fue un instrumento que sirvió para evaluar a niños de Kindergarten y Preparatoria sometidos luego o previamente a otra prueba que intentó ser validada. Los resultados mostraron que las puntuaciones obtenidas por los niños en el MRT tendían a encontrarse en posiciones que permitían que la prueba contara con un techo y un suelo amplio. Como tal, la prueba resulta un buen instrumento para ser normado. En este tipo de instrumentos la medición intenta situar a cada individuo

IX. DISCUSION DE RESULTADOS

Los instrumentos que las maestras de la institución desarrollaron tienen como objetivo evaluar a los alumnos de Kindergarten. Para ello diseñaron dos tablas de contenidos (ver tablas 5 y 6) para la elaboración de la evaluación de la lectura. Esta responde a los programas establecidos por la institución para el área de español. Esta es una institución bilingüe (programa inglés - español) que imparte la instrucción en lectura para ambas lenguas de manera simultánea. El programa intenta responder a destrezas específicas en el área de pre-primaria, aunque actualmente se encuentra en etapa de transición hacia métodos que involucren el aprendizaje cooperativo y mayor integración de las distintas áreas académicas. En otras palabras, el programa parece acercarse más a aquellos aspectos descritos por las corrientes más recientes en oposición a las formas más tradicionales. Sin embargo, tal como se puede observar por la tabla de contenidos, la preocupación de las maestras parece más bien concentrada en aquellas destrezas que podrían considerarse tradicionales.

Las tareas que Gough, Larson & Yopp (1977) enumeran para evaluar las destrezas relacionadas con la conciencia fonológica pueden clasificarse en: discriminación auditiva, incorporación fonética, conteo fonético, eliminación fonética y segmentación fonética. Como puede observarse en el caso de la prueba de medio año de Kindergarten, respecto al área de prelectura se recurrió solamente al conteo fonético (palabra larga-corta) y segmentación fonética (rimas, reconocimiento de sonidos inicial, final, etc.). Las otras tareas se encuentran relacionadas con otros tipos de ejecución, las cuales fueron consideradas por las maestras como necesarias para la adquisición de la lectura: la capacidad de discriminar y manipular los estímulos visuales utilizados para representar sonidos, y el manejo cognoscitivo de éstos. En la evaluación de fin de año puede observarse un fenómeno similar: discriminación auditiva (asociación sonido-símbolo), segmentación fonética (sonido inicial), tareas relacionadas con el manejo de los estímulos visuales utilizados para la representación de los sonidos. Entonces, las

respecto del resto de personas que fueron evaluadas. Con tales resultados se intenta estimar cuanto más o cuanto menos del constructo bajo medición posee el individuo en comparación al resto de personas (Glaser, 1971).

Los resultados de la evaluación desarrollada por la maestra no presentaron de manera tan marcada esta característica. Más bien, reflejaron que niños habían adquirido cierto nivel de rendimiento en tareas relacionadas con el constructo. La prueba diseñada para ser utilizada durante medio año mostró un techo más amplio que la diseñada para evaluar al final del ciclo escolar. Pero en ambos casos el techo de la prueba diseñada por las maestras fue menor que el techo de la prueba MRT correspondiente. En cierto sentido las pruebas elaboradas por las maestras pueden funcionar mejor como un instrumento de criterio, o sea aquel en el que la comparación del individuo no se realiza respecto del resto de personas evaluadas sino respecto de un nivel de ejecución esperado en las tareas relacionadas con un constructo que funciona como criterio (Glaser, 1971).

Las pruebas desarrolladas por las maestras mostraron un buen número de reactivos que aportaron poco en función de discriminar a aquellos niños que se encontraban dentro de un grupo alto respecto de los otros niños del grupo. Pero estos mismos reactivos evaluaban aspectos relacionados con los objetivos de las maestras, como ya se ha mencionado anteriormente. De esta manera, tales reactivos funcionaron como un criterio para distinguir a aquellos alumnos que habían alcanzado el objetivo planteado al iniciar la instrucción.

Un instrumento criterio puede ser normado, debido a que sus contenidos implican muestras de los aspectos relacionados con algún constructo, aunque lo inverso no sea tan sencillo (Popham & Husek, 1971). La validación a través del criterio del MRT permite establecer la correlación que refleja qué similar es el comportamiento de ambas pruebas al medir el constructo. En este caso la evaluación desarrollada para fin de año mostró una correlación alta y satisfactoria con el criterio. La correlación en el caso de la prueba desarrollada para medio año no alcanzó ese mismo nivel. Por otra parte, fue en esta última en que hubo 11 reactivos que parecen resultar fáciles para un buen porcentaje de individuos, aunque no para todos. En este caso los reactivos no funcionaron contrarios a

la orientación esperada con respecto a su funcionamiento dentro de la prueba. Por ello estos reactivos parecen buenas herramientas para establecer un criterio.

De tal manera, que los instrumentos desarrollados por las maestras son instrumentos válidos para la evaluación de los alumnos de Kindergarten. Sin embargo, poseen características que se adecuan mejor a funcionar como un instrumento criterio. Este criterio serían los objetivos establecidos por las maestras y en base a los cuales se orientó la instrucción así como la prueba. Como funcionaron en comparación con el MRT, los instrumentos parecen ser adecuados para evaluar el aprestamiento a la lectura. Cuentan con un beneficio secundario en función de la institución, ya que presentan reactivos que funcionan como criterio para promoción.

Esto es importante, pues como se observó, las pruebas de las maestras contienen los elementos de reconocimiento de las partes que componen una palabra pero no evalúan su manipulación. Así, la evaluación no está completa respecto del constructo aunque satisface los criterios de evaluación para los contenidos planificados por las maestras para el segmento de tiempo sometido a evaluación.

Hasta aquí se han analizado los beneficios que las pruebas desarrolladas por las maestras aportan para verificar el desarrollo académico de los grupos involucrados y para intervenir en los programas. Además, estas pruebas pueden contribuir en la decisión de promoción y retención de los niños si se les conserva como instrumento criterio. Esto no implica que el instrumento podrá determinar potencial para el aprendizaje de la lectura, sino que indicará qué niños no logran un nivel satisfactorio de desempeño en el momento de la evaluación en tareas que la institución consideró relevantes de acuerdo con su programa, tal y como está planteado en los grados que tienen como objetivo principal la instrucción de la lectura.

X. RECOMENDACIONES

Es conveniente mencionar algunos elementos que debieran incorporarse a los instrumentos. Un aspecto que en definitiva debiera tomarse en cuenta es la manipulación de fonemas. El hecho de que las maestras no lo hayan incluido en su instrumento ignora un aspecto importante de la conciencia fonológica la cual, como hemos visto en la "Fundamentación Teórica", ha demostrado ser de gran utilidad como predictor del funcionamiento lector. Tal y como los instrumentos se encuentran en este momento, la capacidad de segmentación e identificación son a los que se les ha proporcionado más peso. Aunque estas medidas tienen mérito y son necesarias, no proporcionan información respecto la capacidad de los alumnos para utilizar la información con flexibilidad al enfrentarse con nuevos textos.

Otras limitaciones que deben mencionarse respecto de las pruebas desarrolladas por las maestras de la institución se encuentran a continuación. En primer lugar, se dedica una buena porción del instrumento a evaluar percepción visual. Este elemento es importante por cuanto aporta información de base respecto a la capacidad del niño para poder distinguir los grafemas. Sin embargo, sacrifica espacio que podría utilizarse para la evaluación de conteo fonético, eliminación fonética y otros tipos de manipulación. Esto debe tomarse en cuenta como limitación en cuanto implica que la prueba no abarca más áreas del constructo de interés. Pero también trae consigo otro dilema. Este es el hecho de que la buena percepción visual no resulta un eficiente predictor del aprendizaje de la lectura, particularmente en comparación a la conciencia fonológica (Gough & Larson, 1997; Gunning, 1992). Aún más, el agregar esta sección al instrumento implica que se está tomando como de alta importancia una variable que la teoría referente a la conciencia fonológica toma como menor. Sin embargo, es importante notar que aunque la buena percepción visual no parece ser un adecuado predictor del aprendizaje de la lectura, problemas en la percepción visual son predictores de problemas en la adquisición de la lectura (Gunning, 1992). De tal manera que el instrumento creado por las maestras

presenta una sección dedicada a discriminar a aquellos alumnos que pudiesen presentar problemas. Debido a que éste no es el propósito del instrumento, y tomando en cuenta que tal área debiese ser evaluada por otra parte, sería aconsejable eliminar la serie dedicada a la percepción visual. El instrumento tiene como objetivo evaluar el aprestamiento a la lectura. Por ello, la percepción visual no es un elemento de alta significancia para este instrumento pues provee información respecto a probables dificultades en el aprendizaje de la lectura sólo si el estudiante obtiene resultados por debajo de los esperados para su edad, mientras que no aporta información cuando estos resultados se encuentran en la norma o sobre ella. Sería más aconsejable crear una evaluación específica para el área de percepción visual. Su aplicación y resultados complementarían los de la evaluación de prelectura, pero sin confundir la información proporcionada por esta última.

Otra de las dificultades que presenta esta prueba es que se desea que su aplicación sea grupal. Esto implica que la evaluación debe ser capaz de distinguir a aquellos niños que pueden realizar las tareas fonéticas sin necesidad de analizar su funcionamiento uno a uno. Esta dificultad hace más complicado el desarrollo de los reactivos. A continuación se dan algunos ejemplos de reactivos que pueden llenar tal propósito.

Para el área de conteo fonético, no es necesario que el estudiante sea capaz de decir que una palabra tiene X número de fonemas. Sin embargo, es importante que pueda discriminar que una palabra tiene más que otra aunque la longitud al pronunciarla no sea notable. Para el desarrollo de un reactivo de este tipo, pueden utilizarse palabras similares, las cuales sólo son diferentes por uno o dos fonemas. Se le pide al niño que marque aquella que es más larga. La diferencia de longitud estará dada sólo por un sonido, lo cual implica que el estudiante es capaz de realizar un "conteo" de los fonemas, o tan sólo notar que hay un sonido más. El propósito del reactivo, sin embargo, sigue intacto, es decir verificar que el estudiante es capaz de percatarse que una palabra tiene más fonemas que la otra. Por ejemplo, se le presentan el dibujo de una casa y el dibujo de una plaza. El evaluador dice ambas palabras señalando los dibujos y pide que los estudiantes marquen cuál de estas es la más larga o la más corta. En este caso, para

países en los que no se hace distinción entre la "s" y la "z" la diferencia se encuentra en que una palabra tiene un fonema más que la otra.

Respecto de la eliminación fonética, se proporciona una palabra de manera oral y se pide a los estudiantes que le quiten uno de los fonemas. Un grupo de dibujos puede servir como distractores y clave, los cuales son nombrados y señalados por el evaluador. El estudiante marca aquella que representa la nueva palabra. Por ejemplo, se les pide a los niños que le quiten la primera letra a "loro". La respuesta a marcar tendría que representar "oro". Lo inverso puede realizarse para evaluar si los alumnos pueden incluir un nuevo fonema a una palabra. Por ejemplo, se pide a los niños que agreguen una "r" al final de la palabra "amo", siendo la respuesta "amor", o en la palabra "vestido" para formar la palabra "vestidor".

Las otras tareas, específicamente las de discriminar y reconocer fonemas iniciales y finales, están ya desarrolladas en los instrumentos de las maestras. Sin embargo, cabe aquí dar algunas sugerencias específicas con respecto a algunos reactivos de los incluidos en las pruebas desarrolladas por las maestras.

Respecto de la prueba de medio año, existe un reactivo que tiene como propósito evaluar percepción visual. El estudiante debe marcar entre una serie de figuras, aquella que es igual a la figura raíz. Este fue uno de los reactivos que produjo resultados menos deseables. La característica particular de este reactivo fue que las formas estaban completamente coloreadas y la figura no contaba con bordes de un distinto color. Es decir, tanto la raíz como las opciones eran manchas negras. El hecho de que la figura fuera una sombra completa, en vez de tener claramente señalada su silueta parece haber actuado desfavorablemente. Los otros reactivos de esta sección funcionaron mejor, por lo que parece conveniente utilizar sólo figuras que muestren la silueta, pero que no estén rellenas del mismo color que las delimita. Otro reactivo muestra en la raíz un helado, y el estudiante debe encontrar entre un grupo otro que se encuentra en la misma posición. Este reactivo parece haber dado algún problema por el tamaño del dibujo, el cual es menor a .33 pulgadas. En este caso se sugiere incrementar su tamaño.

Los reactivos ya mencionados tienen que ver con percepción visual. Ahora es necesario hablar de las destrezas fonéticas, siempre en la prueba de medio año. Una rima que parece haber mostrado ciertas dificultades fue la de "pato-gato". Quizás la dificultad se encuentra en la complejidad de discriminar entre la "g" y la "p". En este caso sería pertinente eliminar esta rima y sustituirla por otra, pues en lugar de ser un reactivo difícil que proveyera de techo a la prueba, parece más bien haber resultado ambiguo.

Respecto de la prueba de fin de año hubo algunas dificultades y ambigüedad en los reactivos relacionados con distinguir el sonido inicial. En este caso los evaluados debían encontrar el dibujo de una palabra que empezaba con el mismo sonido que el nombre del dibujo raíz. Las dificultades parecen haberse relacionado con el uso de vocabulario inusual para los niños, específicamente barril, delantal y torre. Aunque estas palabras, según se consultó con las maestras, forman parte del léxico en algunas lecciones, y la palabra "torre" es hasta cierto punto utilizada en el vocabulario cotidiano, es probable que sean palabras a las que los estudiantes de este nivel no recurren corrientemente. Es de suponer que aquellas palabras a las que recurrimos con naturalidad nos son más familiares, y por ello resultan más fáciles de manipular. En este caso los reactivos parecen haber proporcionado techo estadístico a la prueba, aunque el caso de aparear torre con tomate específicamente parecía ligeramente ambiguo. Sería aconsejable ensayar el uso de este último antes de decidir si se le incluye en una nueva evaluación.

XI. IMPLICACIONES DEL ESTUDIO

En general el funcionamiento de las pruebas hechas por las maestras fue apropiado, y varios de sus reactivos presentaron funcionamiento apropiado. Por ello es recomendable aprovechar sus aspectos positivos enriqueciéndolas con la información obtenida tras cada aplicación.

Por otra parte, las implicaciones del presente estudio para la evaluación en general deben ser analizadas respecto de las poblaciones para las que el instrumento fue desarrollado. En este caso, la población bajo estudio fue la de niños guatemaltecos que asisten a la institución en Kindergarten y Preparatoria. Ahora bien, estos niños cuentan con cierta cantidad de recursos y la proporción de maestro a alumno es tal que los análisis implícitos en el trabajo de evaluación de la conciencia fonológica es posible. Para otras poblaciones, particularmente para aquellas que deben asistir a instituciones que no cuentan con la misma proporción de niños por maestro y con recursos financieros, la evaluación y entrenamiento en destrezas de conciencia fonológica es ciertamente recomendable, aunque rara vez accesible.

Al tomar en cuenta el cambio teórico que se realiza sobre la base de los estudios de conciencia fonológica, es notable cómo los aspectos visomotores tienden a ser más apropiados para el manejo de la información en evaluaciones de niños que ya presentan dificultades en la lectura, por razones directamente relacionadas a este aspecto. Por ello, una consideración que debe tomarse en cuenta, es la forma y organización de los programas de lectura, y particularmente los de pre-lectura. Así como en el instrumento analizado en este trabajo se sugiere que se reconsidere la serie que tiene que ver con la percepción visual, también sería conveniente estudiar si tales cambios pueden ser considerados al tomar decisiones curriculares.

Finalmente, es interesante observar cómo el cambio en la investigación actual refleja un estado de crisis paradigmática tal y como definido por T.S. Kuhn (1970) entre las nuevas tendencias y aquellas que daban mayor trascendencia a la visopercepción. En este caso la presencia de evidencia contradictoria respecto de la percepción visual ha sido

una “anomalía” que ha producido la “crisis” actual. Aunque en otros países la “revolución” de ideas respecto de la lectura pueda encontrarse ya en sus etapas finales, regresando el campo a un estado más o menos estable, en Guatemala pareciera que el conocimiento respecto de esta área continúa en crisis. Es decir, aún existen fuertes seguidores de teorías que apoyan ideas que toman en cuenta las variables visoperceptivas con mayor relevancia que las nuevas tendencias. Este estado de cosas puede ser solucionado sólo a través de un mayor grado de apertura a nuevas ideas. Además, una actitud favorable hacia la experimentación con nuevas ideas, contribuye a que tales concepciones de la realidad puedan demostrar si son válidas o no. En este caso la evaluación y entrenamiento sobre conciencia fonológica es un nuevo elemento en la constelación de principios acerca del aprendizaje de la lectura, uno que debe analizarse a la luz de evidencia concreta.

XII. BIBLIOGRAFÍA

- Byrne, B. & Fielding-Barnsley, R. Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. Journal of Educational Psychology, vol. 85, # 1, 104-111.
1993
- Capper, J. Testing to learn/learning to test. Washington, DC: A.E.D.
1996
- Clancy, C.H. & Pianta, R.C. The Metropolitan Readiness Test as a descriptor and predictor of children's competence in Kindergarten through grade two. Journal of Psychoeducational Assessment, 11, 144-157.
1993
- Crocker, L. & Algina, J. Introduction to Classical & Modern Test Theory. USA: Holt, Rinehart and Winston, Inc.
1986
- Bjaalid, I-K., Høien, T. & Lundber, I. The contribution of orthographic and phonological processes to word reading in young Norwegian readers. Department of Psychology publications.htm. – Explorador de Internet de Microsoft 17-11-97 07:21:46 AM
1994
- Ehri, L.C. Becoming an alphabetic reader is not easy for emergent readers. International Journal of Educational Research, vol. 19, 7, 607-613.
1993
- Flavell, J.H. Metacognitive aspects of problem solving en L.B. Resnick (ed.) The nature of intelligence. New York: John Wiley & Sons.
1976
- Glaser, R. A Criterion-Referenced Test en W. James Popham (ed.) Criterion-referenced measurement: An Introduction. New Jersey: Educational Technology Publications.
1971
- Glaser, R. Instructional Technology and the Measurement of Learning Outcomes: Some questions en W. James Popham (ed.) Criterion-referenced measurement: An Introduction. New Jersey: Educational Technology Publications.
1971
- Goodenough, F.L. Mental Testing. New York: Rinehart & Company, Inc.
1949
- Gough, Ph.B. & Larson, K.C. The Structure of Phonemic Awareness. recife.html – Explorador de Internet de Microsoft 17-11-97 07:27:45 AM
1997
- Gronlund, N.E. How to Make Achievement Tests and Assessments. (5th ed.) Massachusetts: Allyn and Bacon.
1993

- Gunning, Th.G. Creating Reading Instruction for All Children. (2nd ed.) USA: Allyn & Bacon (A Simon & Schuster Company).
1992
- Joyce, B.G. & Wolking, W.D. Standardized tests and timed curriculum-based assessments: A comparison of two methods for screening high-risk students. Journal of Psychoeducational Assessment, 3, 185-193.
1987
- Kuhn, T.S. The structure of scientific revolutions. Chicago: University of Chicago Press.
1970
- Lightbown, P.M. & Spada, N. How languages are learned. Hong Kong: Oxford University Press.
1993
- Nunnally, J.C. & Bernstein, I.J. Teoría Psicométrica. (3a. ed). México: Mc Graw Hill.
1995
- Nurss, J.R. & McGauvran, M.E. Technical Supplement, Metropolitan Readiness Tests, Spanish Edition. USA: The Psychological Corporation-Harcourt Brace Jovanovich, Inc.
1993
- Popham, W. J. & Husek, T.R. Implications of Criterion Referenced Measurement en W. James Popham (ed.) Criterion-referenced measurement: An Introduction. New Jersey: Educational Technology Publications.
1971
- Remmers, H.H., Gage, N.L., Francis-Rummel, J. A practical introduction to measurement and evaluation. (2nd ed.). New York: Harper & Row, Publishers.
1965
- Stahl, S.A., Pagnucco, J.R. & Suttles, C.W. First graders' reading and writing instruction in traditional and process-oriented classes. The Journal of Educational Research, vol. 89, # 3, enero/febrero, 131-144.
1993
- Vellutino, F.R. Introduction to three studies on reading acquisition: Convergent findings on theoretical foundations of code-oriented versus whole-language approaches to reading instruction. Journal of Educational Psychology, vol83, # 4, 437-443.
1991
- Watkins, E.O. & Wiebe, M.J. Factor analysis of the metropolitan readiness test on first-grade children. Psychology in the schools, vol. 21, octubre, 429-431.
1984
- Woolfolk, A.E. Psicología educativa. (3^a ed.) (R.G. Aguilar, Trad.) México: Englewood Cliffs.
1987

APENDICE A

CONCEPTOS Y TERMINOLOGÍA ESTADÍSTICA.

El propósito de esta sección es definir someramente algunos de los términos técnicos y estadísticos utilizados y para aclarar por qué se utilizaron en el presente estudio.

Uno de los conceptos a los que se recurre es el de “confiabilidad de la prueba”, la cual en este caso se mide con el coeficiente alfa. Este hace referencia a la consistencia interna de los datos como una forma de determinar confiabilidad al analizar las interacciones entre reactivos. Determina si la magnitud de los errores de medición que causan que la ejecución varíe de reactivo a reactivo dentro de la misma prueba es significativa o no. De esta forma proporciona una buena estimación de la confiabilidad controlando el muestreo del contenido, el cual es la principal fuente de error de medición para constructos estáticos (Nunnally & Bernstein, 1995). Además, es bastante significativo debido a que es sensible al “muestreo” de factores situacionales, así como de contenido del reactivo (Nunnally & Bernstein, 1995).

Al analizar los reactivos también se toma en cuenta la confiabilidad específica de cada uno de éstos. El índice de confiabilidad del reactivo es una función entre la variabilidad del reactivo y la correlación biserial entre el reactivo y la puntuación total en la prueba. Esto permite pesar la contribución relativa de un reactivo particular a la confiabilidad general de la prueba al tomar en cuenta su varianza (para dos reactivos con correlaciones equivalentes con la puntuación total, aquel con varianza más alta aporta en mayor medida a la confiabilidad interna global). Este dato, sin embargo, presenta ventajas significativas al tomar decisiones acerca de reactivos que se encuentran fuera del rango de dificultad moderada (Crocker & Algina, 1986).

Otro concepto importante que se utiliza es el error de medición. Es apropiado suponer que las medidas obtenidas son en realidad una desviación del valor verdadero (Nunnally & Bernstein, 1995). Esta desviación, o error de medición, puede ser el resultado de un factor o la mezcla de varios de éstos. Estos factores pueden ser tanto

sistemáticos como aleatorios. Cuando el error es sistemático puede, a) afectar a todas las observaciones por igual y ser un error constante, o b) afectar sistemáticamente sólo a ciertos tipos de observaciones y ser un sesgo (Nunnally & Bernstein, 1995). Para ilustrar el concepto puede analizarse el ejemplo citado por Nunnally & Bernstein (1995). Según este ejemplo un termómetro mal calibrado que siempre marca tres grados de más ilustra un error constante. Si el termómetro se encuentra localizado en un lugar tal que cada vez que hay una ráfaga de aire muestra una temperatura 2 grados menor que la temperatura ambiente, la medición está siendo afectada por un atributo irrelevante y por lo tanto está sesgada. Si la medición es afectada por algún atributo irrelevante el error sería un sesgo. Pero el error es aleatorio si la persona que lee el termómetro transpone los dígitos ocasionalmente al registrar las observaciones (Nunnally & Bernstein 1995).

Para el análisis de los datos se recurrió a la correlación. Este es el análisis sugerido por Campbell y Fiske, según reportan Nunnally y Bernstein (1995), para determinar la validez de una medición en base a otra medición ya aceptada como válida. La correlación específica de este estudio es la producto-momento de Pearson. La característica de ésta es que especifica la magnitud de la relación lineal entre dos variables que presentan puntuaciones en distribuciones continuas. Su signo y magnitud denotan la dirección y el grado de relación entre las dos variables. La r , abreviación con que se le designa, no puede ser mayor que +1.00 ni menor que -1.00. Cuando la r es positiva indica que a mayor puntuación de la variable A, también mayor puntuación de la variable B o viceversa. Por el contrario, una r negativa indica que a mayor puntuación de la variable A, menor puntuación de la variable B o viceversa. Mientras más cercana a 1.00, ya sea positivo o negativo, mayor la correlación entre las dos variables (Nunnally y Bernstein, 1995).

También se hace referencia a los términos “techo” y “suelo” de la prueba. Ambos conciernen el rango de niveles de habilidad que la prueba es capaz de abarcar. El “techo” se refiere a la máxima habilidad que el instrumento es capaz de reflejar, mientras que el “suelo” se refiere a la mínima habilidad que le es posible detectar (Goodenough, 1949). Así, una prueba con un techo alto será capaz de detectar aquellos individuos que poseen

un alto nivel de la destreza evaluada. Una prueba con un techo bajo, será capaz de detectar los casos extremos por poseer un nivel muy pobre en la destreza evaluada. Por ejemplo, con una prueba de habilidad numérica con un techo alto y un suelo bajo se podrá detectar tanto a los individuos con francas deficiencias en matemáticas, como a aquellos que cuentan con niveles sobresalientes de la destreza.

Como puede notarse, estos tres tipos de análisis son esenciales para el presente estudio. Aportan la información que nos indica si existe relación o no entre las variables que hemos seleccionado, y además nos informan de cuán confiables son nuestros resultados.

APENDICE B

Análisis de reactivos determinados en base a la muestra de Kindergarten de la institución del MRT-español Nivel 1.

# de Reactivo	Media del reactivo	DS del reactivo	R total del reactivo	Índice de confiabilidad del reactivo
1	0.929	0.258	0.142	0.037
2	0.700	0.458	0.213	0.098
3	0.771	0.420	0.286	0.120
4	0.657	0.475	0.379	0.180
5	0.314	0.464	0.573	0.266
6	0.286	0.452	0.462	0.209
7	0.614	0.487	0.269	0.131
8	0.571	0.495	0.326	0.161
9	0.571	0.495	0.329	0.163
10	0.457	0.498	0.400	0.199
11	0.271	0.445	0.272	0.121
12	0.500	0.500	0.352	0.176
13	0.757	0.429	0.378	0.162
14	0.400	0.490	0.289	0.142
15	0.686	0.464	0.294	0.136
16	0.529	0.499	0.390	0.195
17	0.471	0.499	0.602	0.300
18	0.786	0.410	0.375	0.154
19	0.643	0.479	0.246	0.118
20	0.571	0.495	0.366	0.181
21	0.600	0.490	0.480	0.235
22	0.586	0.493	0.194	0.096
23	0.671	0.470	0.283	0.133
24	0.800	0.400	0.372	0.149
25	0.571	0.495	0.434	0.215
26	0.571	0.495	0.424	0.210
27	0.971	0.167	0.141	0.023
28	0.900	0.300	0.262	0.079
29	0.971	0.167	0.040	0.007
30	0.871	0.335	0.335	0.112
31	0.729	0.445	0.186	0.083
32	0.729	0.445	0.421	0.187
33	0.900	0.300	0.161	0.048
34	0.729	0.445	0.243	0.108
35	0.700	0.458	0.393	0.180
36	0.700	0.458	0.537	0.246
37	0.900	0.300	0.150	0.045
38	0.757	0.429	0.429	0.184
39	0.657	0.475	0.277	0.131

40	0.557	0.497	0.378	0.188
41	0.971	0.167	-0.011	-0.002
42	0.929	0.258	0.064	0.016
43	0.886	0.318	0.168	0.053
44	0.900	0.300	0.144	0.043
45	0.829	0.377	0.538	0.203
46	0.986	0.119	0.226	0.027
47	0.943	0.232	0.079	0.018
48	0.971	0.167	0.302	0.050
49	0.843	0.364	0.063	0.023
50	0.643	0.479	0.214	0.103
51	0.714	0.452	0.314	0.142
52	0.457	0.498	0.157	0.078
53	0.771	0.420	0.210	0.088
54	0.629	0.483	0.418	0.202
55	0.686	0.464	0.240	0.111
56	0.914	0.280	0.131	0.037
57	0.914	0.280	0.251	0.070
58	0.914	0.280	0.431	0.121
59	0.429	0.495	0.120	0.059
60	0.843	0.364	0.326	0.119
61	0.514	0.500	0.382	0.191
62	0.700	0.458	0.415	0.190
63	0.643	0.479	0.418	0.200
64	0.571	0.495	0.292	0.144
65	0.586	0.493	0.310	0.153
66	0.414	0.493	0.281	0.138

APENDICE C

Análisis de reactivos determinados en base a la muestra de Preparatoria de la institución del MRT-español Nivel 2.

# de Reactivo	Media del reactivo	DS del reactivo	R total del reactivo	Índice de confiabilidad del reactivo
1	0.638	0.481	0.139	0.067
2	0.899	0.302	0.346	0.104
3	0.884	0.320	0.414	0.133
4	0.899	0.302	0.176	0.053
5	0.754	0.431	0.295	0.127
6	0.870	0.337	0.069	0.023
7	0.797	0.402	0.283	0.114
8	0.652	0.476	0.148	0.070
9	0.681	0.466	0.377	0.176
10	0.855	0.352	0.125	0.044
11	0.725	0.447	0.326	0.146
12	0.768	0.422	0.099	0.042
13	0.797	0.402	0.490	0.197
14	0.812	0.391	0.235	0.092
15	0.942	0.234	-0.152	-0.036
16	0.928	0.259	-0.110	-0.029
17	0.739	0.439	0.426	0.187
18	0.638	0.481	0.368	0.177
19	0.826	0.379	0.359	0.136
20	0.855	0.352	0.236	0.083
21	0.739	0.439	0.199	0.087
22	0.507	0.500	0.364	0.182
23	0.580	0.494	0.436	0.215
24	0.565	0.496	0.415	0.206
25	0.638	0.481	0.493	0.237
26	0.870	0.337	0.289	0.097
27	0.754	0.431	0.481	0.207
28	0.826	0.379	0.220	0.083
29	0.797	0.402	0.253	0.102
30	0.841	0.366	0.192	0.070
31	0.899	0.302	0.426	0.129
32	0.855	0.352	0.485	0.171
33	0.594	0.491	0.347	0.171
34	0.652	0.476	0.287	0.137
35	0.667	0.471	0.458	0.216
36	0.739	0.439	0.343	0.151
37	0.725	0.447	0.356	0.159
38	0.652	0.476	0.407	0.194
39	0.797	0.402	0.490	0.197
40	0.957	0.204	0.014	0.003

41	0.841	0.366	0.324	0.119
42	0.725	0.447	0.302	0.135
43	0.304	0.460	0.361	0.166
44	0.841	0.366	0.332	0.122
45	0.522	0.500	0.181	0.091
46	0.913	0.282	0.421	0.119
47	0.594	0.491	0.375	0.184
48	0.362	0.481	0.056	0.027
49	0.434	0.499	0.150	0.075
50	0.768	0.422	0.256	0.108
51	0.362	0.481	0.128	0.062
52	0.638	0.481	0.220	0.106
53	0.884	0.320	0.541	0.173
54	0.623	0.485	0.273	0.132
55	0.565	0.496	-0.024	-0.012
56	0.333	0.471	0.186	0.087
57	0.754	0.431	0.236	0.102
58	0.739	0.439	0.082	0.036
59	0.667	0.471	0.317	0.149
60	0.594	0.491	0.166	0.082
61	0.348	0.476	0.061	0.029
62	0.565	0.496	0.077	0.038
63	0.710	0.454	0.134	0.061
64	0.609	0.488	0.477	0.233
65	0.783	0.412	0.194	0.080
66	0.667	0.471	0.509	0.240
67	0.971	0.168	0.011	0.002
68	0.986	0.120	0.109	0.013
69	0.565	0.496	0.326	0.162
70	0.971	0.168	0.191	0.032
71	0.464	0.499	0.349	0.174
72	0.739	0.439	0.124	0.054
73	0.609	0.488	0.289	0.141
74	0.696	0.460	0.196	0.090
75	0.478	0.500	0.350	0.175
76	0.681	0.466	0.454	0.212
77	0.449	0.497	0.219	0.109
78	0.464	0.499	0.458	0.229
79	0.594	0.491	0.369	0.181
80	0.594	0.491	0.350	0.172
81	0.565	0.496	0.405	0.201

APENDICE D

Análisis de reactivos del instrumento elaborado por las maestras para medio año de Kindergarten.

# de Reactivo	Media del reactivo	DS del reactivo	R total del reactivo	Índice de confiabilidad del reactivo
1	0.956	0.205	-0.021	-0.004
2	0.926	0.261	0.096	0.025
3	0.941	0.235	0.333	0.078
4	1.000	0.000	0.000	0.000
5	0.897	0.304	0.097	0.030
6	0.882	0.322	-0.116	-0.038
7	0.971	0.169	0.125	0.021
8	1.000	0.000	0.000	0.000
9	0.691	0.462	0.084	0.039
10	0.735	0.441	0.305	0.135
11	0.868	0.339	0.477	0.162
12	0.868	0.339	0.371	0.126
13	0.985	0.120	0.019	0.002
14	0.868	0.339	0.420	0.142
15	0.912	0.284	0.448	0.127
16	0.632	0.482	0.249	0.120
17	0.544	0.498	0.195	0.097
18	0.471	0.499	0.262	0.131
19	0.515	0.500	0.520	0.260
20	0.559	0.497	0.526	0.261
21	0.971	0.169	0.109	0.018
22	0.868	0.339	0.403	0.137
23	0.838	0.368	0.444	0.163
24	0.765	0.424	0.405	0.172
25	0.941	0.235	0.203	0.048
26	0.809	0.393	0.300	0.118
27	0.632	0.482	0.329	0.159
28	0.706	0.456	0.311	0.142
29	0.985	0.120	0.180	0.022
30	0.971	0.169	-0.138	-0.023
31	0.971	0.169	0.076	0.013
32	0.912	0.284	0.311	0.088
33	0.691	0.462	0.282	0.130
34	0.794	0.404	0.201	0.081
35	0.926	0.261	0.054	0.014
36	0.912	0.284	0.096	0.027
37	0.824	0.381	0.398	0.152
38	0.853	0.354	0.345	0.122
39	0.926	0.261	0.500	0.130
40	0.794	0.381	0.136	0.052

41	0.956	0.205	0.303	0.062
42	0.794	0.404	0.133	0.054
43	0.956	0.205	0.154	0.032
44	0.853	0.354	0.416	0.147
45	0.721	0.449	0.219	0.098
46	0.441	0.497	0.497	0.174
47	0.882	0.322	0.322	0.021
48	0.926	0.261	0.261	0.075
49	0.779	0.415	0.415	0.061
50	0.544	0.498	0.498	0.142
51	0.574	0.495	0.495	0.251
52	0.691	0.462	0.462	0.255
53	0.603	0.489	0.489	0.246
54	0.750	0.433	0.433	0.207

APÉNDICE E

Análisis de reactivos del instrumento elaborado por las maestras para fin de año de Kindergarten.

# de Reactivo	Media del reactivo	DS del reactivo	R total del reactivo	Índice de confiabilidad del reactivo
1	0.954	0.210	0.497	0.104
2	1.000	0.000	0.000	0.000
3	0.969	0.173	0.558	0.096
4	0.985	0.123	0.064	0.008
5	0.969	0.173	0.409	0.071
6	0.938	0.240	0.372	0.090
7	0.938	0.240	0.386	0.093
8	0.938	0.240	0.144	0.035
9	0.985	0.123	0.510	0.063
10	0.985	0.123	0.510	0.063
11	0.985	0.123	0.038	0.005
12	0.815	0.388	0.509	0.197
13	0.892	0.310	0.427	0.132
14	0.831	0.375	0.538	0.219
15	0.923	0.266	0.523	0.139
16	0.754	0.431	0.351	0.151
17	0.954	0.210	0.497	0.104
18	0.662	0.473	0.338	0.160
19	0.523	0.499	0.209	0.104
20	0.877	0.329	0.476	0.156
21	1.000	0.000	0.000	0.000
22	0.800	0.400	-0.011	-0.005
23	0.908	0.289	0.073	0.021
24	0.708	0.455	0.505	0.230
25	0.708	0.455	0.413	0.188
26	0.985	0.123	0.038	0.005
27	1.000	0.000	0.000	0.000
28	0.985	0.123	-0.094	-0.012
29	0.969	0.173	-0.040	-0.007
30	0.969	0.173	0.184	0.032
31	0.892	0.310	-0.031	-0.010
32	0.969	0.173	-0.003	-0.000
33	0.938	0.240	0.104	0.025
34	0.831	0.375	0.144	0.054
35	0.969	0.173	0.184	0.032
36	0.985	0.123	0.011	0.001
37	0.969	0.173	0.016	0.003
38	1.000	0.000	0.000	0.000
39	1.000	0.000	0.000	0.000
40	0.985	0.123	0.011	0.001

41	0.985	0.123	0.116	0.014
42	0.954	0.210	0.066	0.014
43	0.923	0.266	0.014	0.004
44	0.938	0.240	0.225	0.054
45	1.000	0.000	0.000	0.000
46	0.938	0.240	0.332	0.080
47	0.985	0.123	0.274	0.034
48	0.985	0.123	0.142	0.018
49	0.938	0.240	0.131	0.031
50	0.985	0.123	0.274	0.034
51	0.985	0.123	-0.041	-0.005
52	0.985	0.123	-0.041	-0.005
53	0.969	0.173	0.128	0.022
54	1.000	0.000	0.000	0.000
55	0.985	0.123	0.142	0.018
56	0.862	0.345	0.326	0.113
57	0.908	0.289	0.486	0.141
58	0.862	0.345	0.672	0.232
59	0.877	0.329	0.624	0.205
60	0.877	0.329	0.584	0.192
61	0.938	0.240	0.628	0.151
62	0.938	0.240	0.587	0.141
63	0.969	0.173	0.315	0.054
64	0.954	0.210	0.574	0.120
65	0.877	0.329	0.574	0.189
66	0.969	0.173	0.353	0.061
67	1.000	0.000	0.000	0.000
68	1.000	0.000	0.000	0.000
69	0.631	0.483	0.337	0.163
70	0.954	0.210	0.066	0.014
71	0.969	0.173	0.278	0.048
72	0.908	0.289	0.118	0.034
73	0.969	0.173	0.222	0.038
74	1.000	0.000	0.000	0.000
75	0.969	0.173	0.053	0.009