



MINISTERIO DE
EDUCACIÓN PÚBLICA

GOBIERNO
DE COSTA RICA

DGEC
Dirección de Gestión
y Evaluación de la Calidad



MARCO DE REFERENCIA PRUEBA NACIONAL ESTANDARIZADA 2024

Tabla de contenido

Presentación	1
Marco de referencia	2
Propósito y objetivos de la Prueba Nacional Estandarizada	3
Formatos de aplicación	4
Modelo de evaluación	4
Componentes de la prueba	5
Análisis del dominio	6
Establecimiento de los bloques temáticos	6
Establecimiento de afirmaciones	7
Determinación de evidencias	7
Desarrollo de tareas	8
Tipo de enfoque asumido	8
Definición del objeto de evaluación	9
Modelo de medición	10
Teoría de respuesta a los ítems	11
Niveles de desempeño	14
Evidencias de validez y confiabilidad	16
Tipos de ítems en la Prueba Nacional Estandarizada	18
Cantidad de opciones	18
Formatos de ítems de selección de respuesta	19
Aporte de las TIC	21
Diseño Universal de Evaluación	21
Organización de la población estudiantil de acuerdo con su condición asociada o no a discapacidad	22
Organización de la población estudiantil con apoyos educativos por recintos de aplicación de acuerdo con la condición asociada o no a discapacidad	24
Descripción de los apoyos educativos	29
Referencias bibliográficas	32

Presentación

El Decreto Ejecutivo N. ° 43942-MEP considera que la educación es una prioridad para el desarrollo integral del ser humano y el bienestar social, así como el principal instrumento para enfrentar la pobreza, la exclusión y la desigualdad.

Asimismo, al Ministerio de Educación Pública (MEP) como órgano administrador de todo el sistema educativo, ejecutor de los planes, programas y demás determinaciones aprobadas por el Consejo Superior de Educación (CSE) le corresponde promover el desarrollo y consolidación de un sistema educativo de excelencia, el cual permite el acceso de toda la población a una educación de calidad, centrada en el desarrollo integral de las personas y la promoción de una sociedad costarricense que disponga de oportunidades.

Por ello, plantea como imperativo la implementación de un nuevo sistema de evaluación, el cual permita a partir del uso de resultados, la mejora del proceso de enseñanza y aprendizaje congruente con la evaluación de los aprendizajes e integralmente, garantice la calidad del sistema educativo.

En consecuencia, el Consejo Superior de Educación en sesión N° 03-2023, celebrada el 30 de enero de 2023, mediante acuerdo AC-CSE-19-03-2023 y mediante sesión N° 04- 2023, celebrada el 09 de febrero de 2023, mediante acuerdo AC-CSE-30-04-2023, dispuso aprobar la propuesta de modificación al Decreto Ejecutivo N° 40862-MEP de fecha 12 de enero de 2018, denominado Reglamento de Evaluación de los Aprendizajes (REA).

Las modificaciones al Reglamento de Evaluación de los Aprendizajes consideran el diseño y la aplicación de la Prueba Nacional Estandarizada a partir del año 2023.

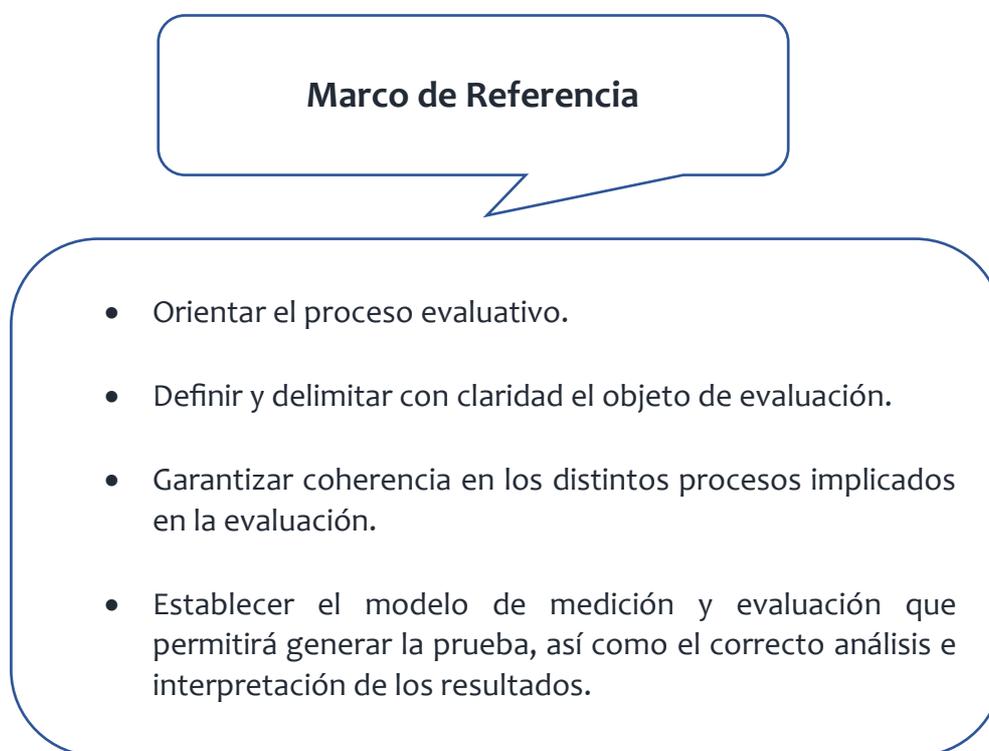
En este documento, presentamos el **Marco de referencia para la Prueba Nacional Estandarizada 2024**, en el que describimos el modelo de evaluación empleado, el cual se ha desarrollado para explicitar con claridad la definición del objeto de medida y para orientar las especificaciones de la evaluación.

Marco de referencia

El marco de referencia para una prueba brinda una noción clara del objeto de evaluación, su extensión y complejidad. En particular, es un requisito técnico y de transparencia en el diseño e implementación de evaluaciones válidas y confiables (Rodríguez Frías & Flotts de los Hoyos, 2019).

En la figura 1 se muestran algunos aspectos que reflejan la utilidad del marco de referencia para la Prueba Nacional Estandarizada.

Figura 1
Aspectos del marco de referencia



Nota. Elaboración propia a partir de Rodríguez Frías y Flotts de los Hoyos (2019).

Propósito y objetivos de la Prueba Nacional Estandarizada

De acuerdo con el artículo 79 del Reglamento de Evaluación de los Aprendizajes, la Prueba Nacional Estandarizada tiene como propósito monitorear el desarrollo de aprendizajes esenciales de los estudiantes para la mejora continua en el proceso de enseñanza y aprendizaje, entre el inicio y final del año académico, asimismo, para la medición de la calidad educativa.

De acuerdo con el artículo 81, son objetivos de la Prueba Nacional Estandarizada:

- a) Contribuir al proceso de mediación pedagógica para la adquisición de habilidades, conocimientos, actitudes y valores en las personas estudiantes.
- b) Coadyuvar en el proceso de enseñanza y aprendizaje orientado al logro de las competencias requeridas para el ciudadano del futuro, que son:
 - Competencias para la ciudadanía responsable y solidaria.
 - Competencias para la vida.
 - Competencias para el empleo digno y el emprendimiento.
- c) Coadyuvar en la determinación promocional efectiva de los estudiantes.
- d) Generar los insumos a partir de los resultados de la prueba para el seguimiento de estrategias de mejora continua del proceso de enseñanza y aprendizaje, de la mediación pedagógica y la evaluación de los aprendizajes.

La Prueba Nacional Estandarizada, evaluará conocimientos, habilidades y otras destrezas propias de las asignaturas según los enfoques y fundamentación teórica de los Programas de Estudio de las asignaturas de Español, Matemáticas, Ciencias (Biología, Física y Química), Estudios Sociales y Educación Cívica (secundaria).

La Prueba Nacional Estandarizada se conforma de dos aplicaciones:

- **Diagnóstica:** Se caracteriza por ser censal y anual. Se aplicará en el primer semestre del curso lectivo y permite conocer el estado de los aprendizajes del estudiantado, con el fin de proporcionar insumos para llevar a cabo procesos de mejora y toma de decisiones.
- **Sumativa:** Se caracteriza por ser censal, anual y además es un requisito para el estudiantado de primaria y secundaria para obtener, respectivamente, el Certificado de conclusión de la Educación General Básica en I y II Ciclo y el Título de Bachiller en Educación Media. Se aplicará en el segundo semestre del curso lectivo.

Formatos de aplicación

La Prueba Nacional Estandarizada tendrá dos formatos de aplicación:

- **Digital:** Se utilizará una plataforma informática (en línea), utilizando computadoras. Se realizará en ambientes controlados y supervisados provistos por los laboratorios de informática educativa de los centros educativos. O bien, otros espacios que la Dirección de Gestión y Evaluación de la Calidad (DGEC) en coordinación con las direcciones regionales de educación establezcan como sedes de aplicación y que cumplan con los requerimientos para fungir como tal.
- **Físico:** Es el formato tradicional de papel y lápiz y se utilizará en aquellos casos que por condiciones del centro educativo sea imposible aplicar en formato digital, por ejemplo; para personas estudiantes matriculadas en Centros de Atención Integral (CAI), personas estudiantes matriculadas en Unidad de Atención Integral (UAI), servicios especializados de salud, o bien, casos fortuitos o de fuerza mayor valorados por la DGEC.

Modelo de evaluación

El diseño centrado en evidencias considera la evaluación como un argumento probatorio, esto es, un argumento que parte de lo que observamos que el estudiantado dice, hace o realiza en unas circunstancias concretas, para inferir lo que saben, pueden hacer o han logrado de forma más general (Mislevy et al., 2003).

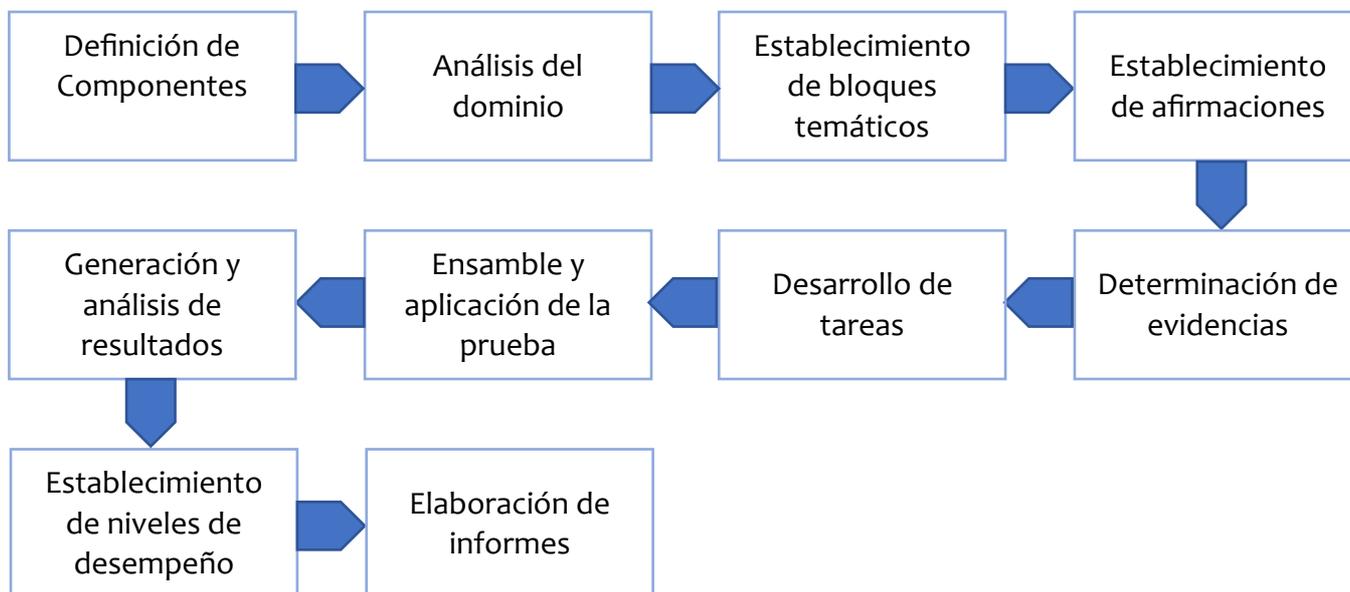
El diseño centrado en evidencias es un marco estructural para analizar y desarrollar evaluaciones como un ejercicio de razonamiento probatorio, con la validez como base para las inferencias que se pretendan hacer a partir de la información recolectada (Mislevy, 2017).

Este modelo se considera un enfoque lógico y sistemático que organiza el trabajo de diseño e implementación de la evaluación en términos de capas o niveles; una metáfora extraída de la arquitectura y la ingeniería de software (Mislevy y Riconscente, 2006). Cada uno de estos niveles cuenta con conceptos clave y un propósito definido que, aunque sugieren un proceso de diseño secuencial, en realidad pueden interactuar, actualizarse e ir mejorando.

En la figura 2 mostramos los pasos vinculados con los niveles del diseño centrado en evidencias, los cuales se pueden interpretar como una cadena de razonamiento en cuanto a argumentaciones y garantías acerca de las inferencias que se puedan hacer sobre los resultados de la población estudiantil en la prueba.

Figura 2

Pasos por seguir de acuerdo con el diseño centrado en evidencias y otros procesos técnicos relacionados al ensamblaje de la Prueba.



A continuación, presentamos la descripción de los niveles adoptados para el diseño, a partir de Mislevy (2017), Mislevy et al. (2003) y Zieky (2014) y otros procesos relacionados con la implementación de la Prueba Nacional Estandarizada 2024.

Componentes de la prueba

Según el Reglamento de Evaluación de los Aprendizajes, la Prueba Nacional Estandarizada se concibe como una sola y no por asignaturas; se centra en la evaluación de conocimientos, habilidades y otras destrezas de acuerdo con los enfoques y la fundamentación teórica de los programas de estudios.

En consecuencia, se establece una relación entre estos elementos y aquellos que son medibles en una prueba de tales características, a los que conocemos como componentes de la prueba y tienen una correspondencia directa con cada una de las asignaturas: Español, Estudios Sociales, Ciencias y Matemáticas (Primaria) y Español, Estudios Sociales, Ciencias, Matemáticas y Educación Cívica (Secundaria).

Análisis del dominio

Se refiere a la recopilación de información sustantiva sobre el dominio que se va a evaluar. En particular, se trata de identificar el conjunto de conocimientos, habilidades y otras destrezas que se pretenden medir, de acuerdo con el propósito de la prueba.

En este nivel, el equipo técnico-asesor de cada componente debe realizar una revisión curricular de la Política Educativa y de la Política Curricular vigentes, así como de la fundamentación de los Programas de Estudio. De esta manera, es posible especificar los conceptos, la terminología, las formas de representación y las maneras de interactuar en relación con los conocimientos, las habilidades y otras destrezas que son objeto de evaluación, luego de haber comprendido su relevancia en la formación del estudiantado, la forma en que son representados en el currículo, la relación entre ellos y la manera en que son adquiridos como parte del proceso educativo formal.

De acuerdo con Mislevy (2006), las investigaciones sobre el aprendizaje también nos dicen mucho sobre cómo las personas llegan a ser competentes en determinados ámbitos y, por tanto, sobre lo que debemos evaluar.

Como primer nivel en el diseño de la evaluación, el análisis del dominio nos permite comprender los conocimientos que las personas estudiantes utilizan en un componente específico, así como las características de las situaciones que evoquen el uso de conocimientos disciplinares, procedimientos y estrategias.

Establecimiento de los bloques temáticos

Un bloque temático se refiere a una agrupación de habilidades y conocimientos que están interrelacionados y que se organizan de manera secuencial con respecto a una temática en particular.

Esta estructura permite que las personas docentes y estudiantes comprendan de una forma adecuada los aprendizajes que se evaluarán en la prueba, ya que se presentan de forma organizada, coherente, con una secuencia lógica y pedagógica, producto de un análisis curricular previo con respecto a los programas de estudio vigentes y políticas educativa y curricular.

Los bloques temáticos toman en cuenta factores como la progresión del aprendizaje y la relevancia de las temáticas. Además, son congruentes con los demás niveles del diseño basado en evidencias, modelo de evaluación asumido para la Prueba.

Establecimiento de afirmaciones

De acuerdo con Zieky (2014), el diseño basado en evidencias se puede considerar como un medio para construir una cadena de argumentos y garantías que respaldan las afirmaciones que se hacen sobre las personas estudiantes, a partir de una prueba.

Las afirmaciones son enunciados acerca de lo que las personas estudiantes muestran en su desempeño en la prueba, por lo que deben ser claras y precisas, así como deben apoyarse en otras afirmaciones que tengan diferentes niveles de especificidad.

El propósito de una prueba y las afirmaciones que se pueden hacer sobre el estudiantado están estrechamente relacionados. En otras palabras, el propósito de la prueba se puede expresar como la afirmación de mayor nivel que se puede hacer a partir de los resultados.

Determinación de evidencias

Las evidencias se basan en los comportamientos o productos observables que se pueden concretar en las respuestas a una tarea específica. Para ello, se debe hacer una descripción detallada de cuáles son los conocimientos y las habilidades en los que se centra la medición.

En este modelo de evaluación, las evidencias brindan las garantías que permiten apoyar o fundamentar las afirmaciones que se han establecido.

Claramente, las evidencias se enfocan en aspectos relevantes de los comportamientos o productos observables, así como la diferenciación de lo que es medible con respecto a lo que no lo es, en una prueba con sus diferentes formatos de aplicación.

Para determinar evidencias se puede partir de una situación ideal o hipotética, no obstante, las restricciones o condiciones reales de la prueba (por ejemplo, formato de aplicación, tiempo reglamentario, entre otros), así como la relevancia para respaldar una afirmación son aspectos cruciales para emprender la siguiente etapa.

En el ejemplo 1 se muestra una afirmación del componente de Ciencias para primaria y sus respectivas evidencias, elaboración que se realizó a partir del Programa de Estudio correspondiente.

Ejemplo 1

Ejemplo de afirmación y evidencias en el componente “Ciencias” para primaria

Afirmación	Evidencias
La persona estudiante distingue los factores que determinan la biodiversidad de Costa Rica y su importancia.	Identifica los factores que permiten una alta biodiversidad en Costa Rica.
	Relaciona la presencia de los diferentes factores con el alto número de hábitats del territorio nacional.

Desarrollo de tareas

El desarrollo de tareas se basa en una descripción de lo que la prueba incluirá para la medición específica del objeto de evaluación. En particular, se enfocará en el tipo de ítems que son situaciones que provocan el comportamiento o producto observable tal y como fue descrito en las evidencias (Zieky, 2014).

En la sección correspondiente se detallará el tipo de ítems que conforma la Prueba Nacional Estandarizada, así como otras características referentes a la cantidad de opciones.

Tipo de enfoque asumido

Como parte del modelo de evaluación utilizado, la Prueba Nacional Estandarizada estará bajo el enfoque criterial para la formulación de juicios de valor. De acuerdo con Ravela (2006), este enfoque consiste en “privilegiar la comparación del desempeño de un individuo con una definición clara y precisa de lo que se espera que conozca y sea capaz de hacer en un determinado dominio” (p. 43).

Las pruebas criterioles se diseñan para brindar una medida del desempeño que es interpretable en términos de un dominio de tareas de aprendizaje claramente definido y delimitado (Ravela, 2006).

Este enfoque, junto al modelo de medición que detallaremos más adelante, nos permitirá informar acerca de la probabilidad de las personas estudiantes de responder ítems de cierto grado de dificultad.

Definición del objeto de evaluación

El objeto de evaluación de la Prueba Nacional Estandarizada está constituido por el conjunto de conocimientos, habilidades y otras destrezas de la persona estudiante para realizar eficazmente tareas vinculadas a las competencias curriculares plasmadas en los Programas de Estudio.

En consonancia con los planteamientos de la Política Curricular acerca de una evaluación transformadora que “se asuma como una forma de identificar la complejidad de los retos y los nuevos elementos que se integran a los nuevos aprendizajes” (MEP, 2015, p. 27), se cuenta con una única prueba integrada por componentes que corresponden a los conocimientos disciplinares vinculados a las competencias curriculares.

De esta manera, la Prueba Nacional Estandarizada no busca medir contenidos específicos ni enciclopédicos, sino que toma como punto de partida los conocimientos previos del estudiantado para desplegar o activar procesos cognitivos presentes en las siguientes habilidades de la dimensión “maneras de pensar” de la Política Curricular:

- **Pensamiento sistémico:** Habilidad para ver el todo y las partes, así como las conexiones que permiten la construcción de sentido de acuerdo con el contexto.
- **Pensamiento crítico:** Habilidad para mejorar la calidad del pensamiento y apropiarse de las estructuras cognitivas aceptadas universalmente (claridad, exactitud, precisión, relevancia, profundidad, importancia).
- **Resolución de problemas:** Habilidad de plantear y analizar problemas para generar alternativas de soluciones eficaces y viables.

De acuerdo con el modelo de evaluación asumido, el análisis del dominio permite establecer una relación entre los componentes, de tal manera que la prueba funcione como un argumento basado en evidencias. Es decir, el dominio está integrado por los componentes que permiten generar afirmaciones de lo que el estudiantado debe saber hacer, para lo que se requiere obtener evidencias cuya concreción se da en tareas particulares (ítems) que son la garantía de las afirmaciones planteadas.

Modelo de medición

Los fundamentos de la Teoría Clásica de los Test (TCT) se establecieron en la primera década del siglo XX, a partir de los planteamientos de Spearman con su modelo lineal clásico, el cual asume que la puntuación empírica (X) de una persona en una prueba consta de dos componentes: la puntuación verdadera de dicha persona (V) y el error (e) asociado a todo proceso de medición. De esta manera, el modelo matemático de esta teoría está dado por la ecuación: $X = V + e$.

De acuerdo con Martínez, Hernández y Hernández (2006) es posible cuantificar las propiedades de los ítems de puntuación dicotómica mediante algunos índices propios de la TCT y que se pueden clasificar en los que:

- 1) Describen la distribución de respuestas a los ítems y que influyen en la media y la varianza de la prueba, por ejemplo, el *índice de dificultad del ítem*, definido como la proporción de personas que lo aciertan. Los valores de este índice oscilan entre 0 y 1; valores muy altos representan ítems muy fáciles, mientras que valores muy bajos representan ítems muy difíciles.
- 2) Describen el grado de relación entre la respuesta al ítem y algún criterio interno o externo a la prueba, por ejemplo, el *coeficiente de correlación biserial-puntual*, según el cual un ítem es eficaz en la medida en que las personas que puntúan alto en el criterio de interés tiendan a puntuar alto en el ítem y las personas que puntúan bajo en el criterio, obtengan puntuaciones bajas en el ítem. Por lo general, tal criterio se basa en la puntuación total de la prueba con el fin de enfatizar en la homogeneidad del conjunto de ítems que constituye el test. El valor de este coeficiente es bastante sensible a los valores del índice de dificultad, por lo que proporciona resultados sesgados a medida que su valor se aproxima a los extremos (-1 y 1) y así favorece a los ítems de dificultad media.
- 3) Son función simultáneamente de la varianza del ítem y de su índice de discriminación, por ejemplo, el *coeficiente α de Cronbach*, el cual refleja el grado en que covarían los ítems que constituyen el test y por ello es un indicador de la consistencia interna de la prueba. Habitualmente se le considera para optimizar la prueba como instrumento de medida, ya que brinda información para elegir los ítems que contribuyen a mejorar el test (Muñiz, Fidalgo, García-Cueto, Martínez & Moreno, 2005).

El surgimiento de la TCT tuvo como objetivo central “encontrar un modelo estadístico que fundamentase adecuadamente las puntuaciones de los tests y permitiera la estimación de los errores de medida asociados a todo proceso de medición” (Muñiz, 2002, p. 19). No obstante, su popularidad se debe en parte a su sencillez matemática y a que sus supuestos básicos son mínimos e incluso débiles, por lo que pueden ser aplicados a diversos conjuntos de datos.

Martínez (2005) enumera cuatro limitaciones principales: 1) los estadísticos del ítem dependen de la tendencia central y de la variabilidad del rasgo en la muestra y en la población, 2) los estadísticos

del test dependen de la muestra y la población en que se calculen, 3) la puntuación empírica total depende de los ítems que componen el test, y 4) no es posible determinar ni especificar el comportamiento de una persona cuando se enfrenta a un ítem.

Teoría de respuesta a los ítems

La Teoría de respuesta a los ítems (TRI) se basa en dos postulados esenciales: 1) el rendimiento de una persona en un ítem se puede predecir o explicar por un conjunto de factores llamados *rasgos* o *aptitudes latentes* [latent trait]; y 2) la relación entre tal rendimiento y el conjunto de rasgos se puede describir por una función estrictamente creciente llamada *función característica del ítem* o *curva característica del ítem*, la cual especifica que a medida que el nivel del rasgo aumenta, la probabilidad de obtener la respuesta correcta para el ítem también aumenta (Hambleton & Cook, 1977; Hambleton, Swaminathan & Rogers, 1991). Estos rasgos suelen denominarse *aptitud* y se simbolizan con la letra θ (Martínez, 2005).

Las estimaciones de las puntuaciones de las personas en los rasgos latentes se utilizan para explicar la puntuación que tendrá la persona en un ítem o en una prueba completa, por lo que la variable independiente es el atributo o rasgo y la variable dependiente es la respuesta al ítem o a la prueba. De esta manera, a diferencia de la TCT, las puntuaciones empíricas son el resultado y no el fundamento del atributo (Martínez, 2005).

Los supuestos de los modelos de TRI son fundamentalmente:

- 1) *Dimensionalidad*, esto es, hay un conjunto de k rasgos responsables del desempeño de una persona en una prueba determinada y que definen un espacio latente de k dimensiones (Martínez, Hernández & Hernández, 2006). En la práctica común se supone la unidimensionalidad, es decir, que una única aptitud o rasgo es suficiente para explicar los resultados de las personas y las relaciones entre los ítems. Hambleton, Swaminathan y Rogers (1991) afirman que este requerimiento consiste en que se detecte la presencia de un componente o factor dominante que influya en el rendimiento de la prueba. Asimismo, tales autores reconocen la existencia de otros factores adicionales que afectan dicho rendimiento, por ejemplo, las habilidades cognitivas de las personas.
- 2) *Independencia local*, es decir, si se mantiene constante la medida del rasgo que explica el desempeño en una prueba, las respuestas de las personas a un par de ítems cualquiera son estadísticamente independientes, o bien, las habilidades especificadas en el modelo son los únicos factores que influyen en las respuestas de las personas a los ítems (Hambleton, Swaminathan & Rogers, 1991).

Cuando se cumple el supuesto de unidimensionalidad, se obtiene la independencia local, por lo que se afirma que ambos supuestos son equivalentes en términos generales.

Según Embretson (1999) hay cinco ventajas fundamentales en la utilización de modelos de TRI, en contraste con la TCT: 1) el significado de los niveles del rasgo aplica para cualquier ítem con

dificultad conocida; 2) las propiedades de los ítems y los niveles de rasgo brindan un modelo completo de comportamiento, pues se incluyen parámetros por separado para personas y para ítems; 3) la objetividad específica, esto es, el nivel de rasgo y las propiedades de los ítems son variables independientes que se pueden estimar por separado; 4) las estimaciones del nivel de rasgo para las personas están controladas por las propiedades de los ítems que fueron administrados; y 5) las estimaciones de la dificultad de un ítem están controladas por los niveles de rasgo de las personas en particular de la muestra que se utiliza en la calibración de los ítems.

La propiedad de invarianza de los parámetros del ítem y de la aptitud es el aspecto central de la TRI y su principal diferencia con respecto a la TCT. De acuerdo con Martínez (2005) esta propiedad supone que los parámetros que caracterizan al ítem no dependen de la distribución de la aptitud de las personas y que los parámetros que caracterizan a la persona no dependen del conjunto particular de ítems que se utilicen para su medición.

Por ello, cuando un modelo de TRI se ajusta a los datos, se obtiene la misma curva para un ítem de una prueba independientemente de la distribución de la aptitud en el grupo de personas usado para estimar los parámetros del ítem. No obstante, en la teoría clásica, la estimación de la dificultad de un mismo ítem para dos grupos que tienen diferentes niveles de habilidad resultará en valores distintos para tales grupos.

De tales postulados básicos es posible notar las ventajas que tiene la TRI sobre la TCT. Por ejemplo, dado el mismo ítem de la misma prueba para dos grupos que tienen diferentes niveles de habilidad, la estimación de la dificultad del ítem será diferente si se basa en la TCT, pero será muy similar si se basa en la TRI.

Por otra parte, en el marco de la TCT, las puntuaciones de las personas son una función de la dificultad de los ítems que se administran, por lo que las comparaciones entre las puntuaciones de las personas en versiones diferentes de la prueba se limitan a las formas paralelas del test, es decir, las comparaciones de las puntuaciones dependen de la prueba. Por el contrario, los modelos de TRI pueden proporcionar las estimaciones más apropiadas para la variable latente basándose en las propiedades de los ítems, aunque a las personas les haya correspondido versiones diferentes de la prueba.

Lo anterior se fundamenta en el hecho de que si se dispone de un conjunto de ítems de los cuales se conocen previamente las estimaciones de sus parámetros y si se tiene un modelo de TRI que se ajuste a los datos, entonces es indiferente qué subconjunto de ellos se utilice, pues todos darán las mismas estimaciones de θ para las personas, es decir, el instrumento utilizado es invariante respecto de la medida de θ (Muñiz, 1997).

En lo que se refiere al error estándar de medición, la TCT supone que es una propiedad de la prueba y, por lo tanto, asume que dicho error es el mismo para todas las personas, es decir, no se toman en cuenta las diferencias en cuanto a sus niveles de aptitud. La TRI ofrece un error de medición para cada persona en cada uno de los diferentes niveles de habilidad, es decir, dado determinado nivel de θ , la *información de un ítem* cuantifica la precisión con que dicho ítem contribuye a estimar ese valor particular de θ .

De lo anterior, a diferencia de la TCT, en la TRI es posible conocer cuáles ítems son los mejores para medir un determinado nivel de habilidad. Así, la *función de información* se define como el recíproco de la precisión con la que se puede estimar un parámetro (la varianza de los errores de medida). Asimismo, la *información de la prueba* para determinado nivel de θ es el resultado de la suma de la información de los ítems que lo componen a la hora de estimar θ (Muñiz, Fidalgo, García-Cueto, Martínez & Moreno, 2005).

Finalmente, las puntuaciones obtenidas a partir de la TCT no tienen significado alguno en términos de cuáles ítems puede o no hacer una persona, ya que los parámetros de los ítems dependen del grupo de personas y la puntuación de cada una depende del conjunto particular de ítems administrados. No obstante, existen modelos de TRI que permiten comparar la estimación de la habilidad de las personas con la dificultad de los ítems al colocarlos en una escala común (en particular, el modelo de Rasch). De esta manera, las estimaciones de habilidad no solo están representadas por las puntuaciones totales o el valor de θ , sino también para diferentes niveles de dificultad de los ítems (Embretson & Reise, 2000).

Dado el tipo de ítems de la Prueba Nacional Estandarizada, en particular, ítems de selección de respuesta, se trabajará con el modelo de TRI para puntuaciones dicotómicas conocido como modelo de Rasch.

El modelo de Rasch brinda aproximaciones de gran utilidad para las mediciones, con lo que es posible entender los procesos que subyacen a la razón por la que tanto las personas como los ítems se comportan de una manera determinada (Bond & Fox, 2001). Su formulación fue planteada por Rasch en el año 1960, quien postuló que la respuesta a un ítem depende solo de la aptitud de la persona y de la dificultad del ítem.

Este modelo proporciona un marco matemático con el cual quienes elaboran pruebas pueden comparar sus datos. Para ello se basa en la idea de medir solo un atributo o rasgo a la vez (unidimensionalidad) en una línea jerárquica para efectos de investigación (en términos de “más que” o “menos que”); esta última es una idealización teórica contra la cual se pueden comparar patrones de respuesta que no coincidan con el ideal o esperado. Las desviaciones de las personas o del rendimiento en los ítems con respecto a dicha línea, alertan a quienes investigan para que reconsideren el planteamiento del ítem o incluso las interpretaciones que hagan de las puntuaciones a partir de esos datos. Tanto la dificultad del ítem como la aptitud de la persona se estiman en una misma escala y cada una de esas estimaciones tiene un grado de error asociado a ellas, el cual decrece a medida que la dificultad y la aptitud aumentan (Bond & Fox, 2001). A esta propiedad se le conoce como *medición conjunta*.

En pruebas de rendimiento, la ubicación de un ítem se conoce como la *dificultad de un ítem*. En general, los ítems que se consideran fáciles son aquellos en los que las personas menos competentes muestran una tendencia a responderlos correctamente y les corresponden valores negativos de θ . Por el contrario, los ítems más difíciles son los que las personas más competentes tienden a acertarlos; a ellas les corresponden valores positivos de θ . Finalmente, los ítems alrededor de 0 se consideran como ítems de dificultad promedio (de Ayala, 2009).

A pesar de que en la mayoría de la literatura sobre modelos de TRI se consideran el modelo de Rasch y el modelo logístico de un parámetro como equivalentes, existen diferencias conceptuales entre ambos y se detallan en la publicación *Rasch Measurement Transactions* del Institute for Objective Measurement (2005); algunas de ellas se describen a continuación.

En cuanto a la probabilidad para datos binarios, el modelo de Rasch plantea la probabilidad de que una persona n que tenga el requisito deseado consiga el éxito al encontrarse con el ítem i . Por otra parte, el modelo logístico de un parámetro plantea la probabilidad global de éxito en el ítem i según la distribución de θ . En lo que se refiere a la evaluación del ajuste, el modelo de Rasch plantea un ajuste de los datos al modelo, es decir, es una valoración local; por el contrario, el modelo de un parámetro es global en el sentido que evalúa el ajuste del modelo a los datos y, en caso negativo, lo que se rechaza es el modelo propuesto.

Otra característica de interés que diferencia ambos modelos es la consideración de la discriminación del ítem. En el modelo de Rasch, la curva característica del ítem se modela para que sea paralela a una recta de pendiente igual a 1 (la ojiva logística natural), mientras que en el modelo de un parámetro resulta paralela a una recta de pendiente igual a 1,7 (aproximándose a la pendiente de la ojiva normal acumulada).

Niveles de desempeño

Los niveles de desempeño corresponden a categorías que describen aquello que es capaz de hacer el estudiantado frente a lo que mide la prueba. Estos niveles son esenciales para interpretar los resultados de la prueba y tomar decisiones por parte de los diferentes actores del sistema educativo.

Estos niveles tienen una lógica jerárquica; en otras palabras, un estudiante que se ubica en un nivel tiene altas probabilidades de resolver correctamente los ítems que corresponden a ese nivel y los de niveles inferiores.

Para la Prueba Nacional Estandarizada se definen tres niveles de desempeño: básico, intermedio y avanzado. Las descripciones generales de cada uno de los niveles, se muestra a continuación:

Nivel	Descripción
Básico	Corresponde a un dominio elemental de conocimientos, habilidades y destrezas escolares incluidas en los componentes evaluados. Denota un desempeño estudiantil poco satisfactorio; no obstante, brinda insumos para elaborar estrategias las cuales guíen hacia el mejoramiento que posibilite avanzar, gradualmente, al nivel intermedio.

Nivel	Descripción
Intermedio	Corresponde a un dominio parcial de conocimientos, habilidades y destrezas escolares incluidas en los componentes evaluados. Denota un desempeño estudiantil en proceso para alcanzar el dominio; además, brinda insumos para elaborar estrategias las cuales guíen hacia el mejoramiento que posibilite continuar progresando, gradualmente, al nivel avanzado.
Avanzado	Corresponde a un dominio satisfactorio de conocimientos, habilidades y destrezas escolares incluidas en los componentes evaluados. Evidencia una alta probabilidad de que el desempeño estudiantil esté en concordancia con las afirmaciones propias del nivel que permitan profundizar sobre los saberes y habilidades de forma significativa.

Los niveles de desempeño de la Prueba Nacional Estandarizada son establecidos utilizando un proceso riguroso que incluya la participación de asesores pedagógicos a nivel nacional, lo cual aporta evidencias de validez al proceso. El establecimiento de cada uno de los niveles de desempeño conlleva una serie de procedimientos y actividades que derivan en el estableciendo de puntos de corte. De acuerdo con Herrera et al. (2009) estos puntos son las operacionalizaciones numéricas de las descripciones cualitativas de los niveles de desempeño.

El método utilizado para el establecimiento de puntos de corte para la Prueba Nacional Estandarizada está basado en el método *Bookmark*. Este método considera el juicio experto de un grupo de panelistas, los cuales revisan cuidadosamente los ítems de cada componente de la prueba. A cada uno de los panelistas se le brindan los ítems ordenados de forma creciente por dificultad y en varias rondas de trabajo, cada experto coloca los marcadores que determinan los puntos de corte.

En cada ronda se busca el consenso entre los distintos panelistas, mediante una discusión detallada sobre qué mide el ítem en cuestión y por qué tiene una dificultad mayor que el ítem anterior. Como producto final de estas rondas, se establecen los puntos de corte que permiten clasificar los desempeños en los tres niveles establecidos.

Una de las principales características de este método es que es congruente con el método de medición y evaluación adoptado para la prueba. Además, los niveles de desempeño establecidos con esta metodología permiten una comunicación eficiente de los resultados educativos y la toma de decisiones para el mejoramiento de la calidad educativa

La utilización del método *Bookmark* aporta evidencias de validez a la interpretación de resultados de la Prueba Nacional Estandarizada y permite a los distintos interesados contar con información válida y confiable para la toma de decisiones, en pro de la mejora de la calidad educativa del sistema educativo costarricense.

Evidencias de validez y confiabilidad

De acuerdo con Ravela (2006), el concepto de validez refiere al grado en que los juicios de valor que se formulan de los resultados de una evaluación estén debidamente sustentados y estén efectivamente relacionados con el “referente” definido para la evaluación; mientras que la confiabilidad se refiere a la precisión de las medidas y de la evidencia empírica empleada en la evaluación.

Como parte de las evidencias de validez de la Prueba Nacional Estandarizada, posteriormente al análisis del dominio, se elabora una tabla de especificaciones, la cual operacionaliza el constructo y orienta con claridad el proceso evaluativo. Estas tablas fueron revisadas por asesores regionales del Ministerio de Educación Pública y pueden ser consultadas en el **marco de especificaciones** de la Prueba Nacional Estandarizada.

Otro elemento que aporta evidencias de validez es el estudio piloto, el cual permite obtener datos psicométricos de los ítems para el ensamblaje de las distintas pruebas durante el año 2024.

Además, se cuenta con procesos internos y externos de juzgamiento de ítems por parte de expertos de cada uno de los componentes de la prueba. Las observaciones emanadas en los procesos de juzgamiento son posteriormente valoradas por los equipos técnicos para la incorporación en las versiones finales de los ítems.

En la figura 3, se muestra la tabla utilizada para el juzgamiento de ítems que se le brinda a los equipos de asesores técnicos (nacionales y regionales) con el fin de registrar las evidencias y observaciones al respecto.

Figura 3

Tabla para el juzgamiento de ítems de cada componente de la Prueba Nacional Estandarizada

**PRUEBA NACIONAL ESTANDARIZADA
JUZGAMIENTO DE LA CALIDAD TÉCNICA DEL ÍTEM**

Nombre del juez: _____ **Fecha:** ____/____/____

Habilidades de la Política Curricular					
Pensamiento sistémico <input type="radio"/>		Pensamiento crítico <input type="radio"/>		Resolución de problemas <input type="radio"/>	
Bloque de conocimientos medido	Afirmación medida	Evidencias	Respuesta correcta (clave)		
N° _____	N° _____	N° _____	A <input type="radio"/>	B <input type="radio"/>	C <input type="radio"/>
Contenido			SÍ	NO	
Uso correcto de la gramática y la ortografía.			<input type="radio"/>	<input type="radio"/>	
La redacción del enunciado es coherente y afirmativa.			<input type="radio"/>	<input type="radio"/>	
El vocabulario es adecuado al desarrollo cognitivo del estudiantado.			<input type="radio"/>	<input type="radio"/>	
La información es relevante para el planteamiento y la solución del ítem.			<input type="radio"/>	<input type="radio"/>	
Opciones de respuesta			SÍ	NO	
La clave es incuestionable.			<input type="radio"/>	<input type="radio"/>	
Las opciones de respuesta son excluyentes entre sí.			<input type="radio"/>	<input type="radio"/>	
La redacción de las opciones de respuesta es afirmativa.			<input type="radio"/>	<input type="radio"/>	
Existe concordancia entre el enunciado y las opciones de respuesta.			<input type="radio"/>	<input type="radio"/>	
Sensibilidad					
El ítem presenta un					
lenguaje no inclusivo <input type="radio"/>		vocabulario técnico no pertinente al componente <input type="radio"/>			
El ítem presenta implicaciones		SÍ <input type="radio"/>		NO <input type="radio"/>	
políticas <input type="radio"/>	de género <input type="radio"/>	ideológicas <input type="radio"/>	xenofóbicas <input type="radio"/>	propagandistas <input type="radio"/>	
religiosas <input type="radio"/>	deportivas <input type="radio"/>	etnofóbicas <input type="radio"/>	homofóbicas <input type="radio"/>		
Imágenes*	SÍ <input type="radio"/>		NO <input type="radio"/>		
			SÍ	NO	
La resolución de la imagen es óptima.			<input type="radio"/>	<input type="radio"/>	
La imagen es pertinente para la solución del ítem.			<input type="radio"/>	<input type="radio"/>	
La calidad de la imagen es clara (color, luz y texturas).			<input type="radio"/>	<input type="radio"/>	
La imagen complementa la información del enunciado.			<input type="radio"/>	<input type="radio"/>	

*Imágenes: símbolos, gráficos, dibujos, esquemas, figuras, otros.

Tipos de ítems en la Prueba Nacional Estandarizada

La Prueba Nacional Estandarizada solamente incluye ítems de selección de respuesta, conocidos comúnmente como ítems de selección única, en los que la respuesta correcta se denomina “clave” y las otras opciones se conocen como “distractores”.

En particular, los ítems de selección de respuesta son muy útiles en la medición de conocimientos que requieran distintas demandas cognitivas, aunque claramente no son los únicos existentes. A pesar de que su elaboración no es sencilla, permiten dicotomizar la respuesta en términos de acierto/no acierto para efectos de puntuación en una prueba, lo que facilita la automatización en los procesos de calificación, sobre todo en aplicaciones de gran escala.

Es común que se les asocie con ítems que se respondan tan solo evocando algún término o concepto, sin embargo, realmente permiten medir la aplicación de conocimientos, habilidades y otras destrezas en la resolución de tareas complejas (Haladyna & Rodríguez, 2013).

Cantidad de opciones

La Prueba Nacional Estandarizada está compuesta por ítems de selección única de tres opciones: una opción correcta y dos distractores plausibles.

Al respecto, se toma como base el metaanálisis elaborado por Rodríguez (2005) acerca de la cantidad de opciones considerada como óptima para la construcción de ítems de selección única. En su artículo deja ver cómo, a finales de la década de los noventa, las recomendaciones versaban acerca de redactar tantas opciones como fuera posible o desarrollar tantos distractores funcionales como se pudiera. Asimismo, se planteó que el punto central no era la cantidad de distractores, sino su calidad.

Autores como Haladyna, Downing y Rodríguez (2002) concluyeron que “tres opciones son suficientes en la mayoría de los casos” (p. 318), tomando en cuenta que, en algunas ocasiones, se dificulta que una cuarta opción resulte plausible.

Ante la preocupación de que exista una mayor probabilidad de acertar ítems de tres opciones que aquellos que tengan cuatro o cinco alternativas, Rodríguez (2005) señala que es poco probable que las personas examinadas se involucren en una escogencia sin criterio alguno, sino que eliminarán los distractores menos plausibles en cada ítem. Además, la dificultad de un ítem no solo está en función de la cantidad de distractores utilizados.

Formatos de ítems de selección de respuesta

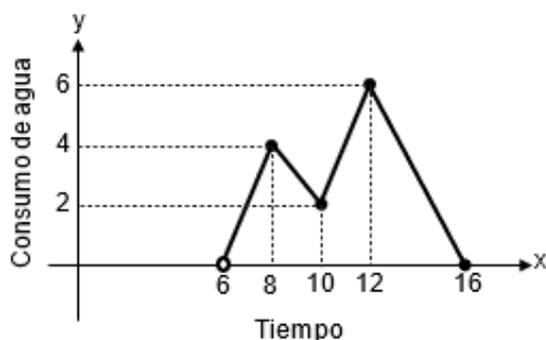
La elaboración de ítems requiere de personas expertas en el área de conocimientos por medir, así como experiencia en el ámbito educativo. Este es un proceso que también involucra creatividad y claridad en cuáles son las habilidades y otras destrezas que se emplean en la resolución de tareas de distinta índole y complejidad.

Los ítems de selección de respuesta pueden presentarse de forma individual o como multirreactivos; este último formato permite evaluar diversos conocimientos y habilidades interrelacionados que se desprenden de la información que se muestra al inicio de los ítems. Cada pregunta se resuelve y evalúa de manera independiente.

En el ejemplo 2, presentamos un ejemplo de multirreactivo en el componente “Matemáticas” para secundaria.

Para responder los ítems 12 y 13 considere la siguiente información:

La siguiente representación gráfica corresponde a la cantidad de agua potable, en metros cúbicos, que se consumió en una institución en función del tiempo “ x ”, en horas de un día, con $6 < x \leq 16$:



12) La cantidad de agua que se consumió, en esa institución, disminuyó entre las

- A) 6 h y 8 h.
- B) 8 h y 10 h.
- C) 10 h y 12 h.

13) ¿A qué hora de ese día se consumió la mayor cantidad de agua en esa institución?

- A) A las 6 h
- B) A las 12 h
- C) A las 16 h

Aunque existen diversos formatos para los ítems de selección de respuesta, en la Prueba Nacional Estandarizada se utilizarán los dos siguientes:

- **Formato interrogativo:** Se presenta una situación y una pregunta con base en ella. Permiten medir tanto la comprensión como la aplicación de conocimientos, habilidades y otras destrezas para seleccionar la respuesta correcta.
- **Formato de encabezado incompleto:** Se presenta, en el encabezado, una oración incompleta; las opciones completan el encabezado gramaticalmente (debe existir concordancia entre género y número).

Asimismo, en la Prueba Nacional Estandarizada se emplearán algunas formas complejas de ítems de selección de respuesta expuestos en Aiken (2003) y en Haladyna y Rodríguez (2013). En el ejemplo 3, mostramos un ítem de selección de respuesta en que las opciones presentan una justificación para la respuesta.

En el ejemplo 3 presentamos un ejemplo de ítem con opciones que presentan justificación para la respuesta

Una maestra recolectó datos, por medio de una encuesta, sobre el número de horas diarias que dedican al estudio los 30 estudiantes de una sección. A continuación, se muestra el gráfico que elaboró la maestra a partir de los datos recolectados:



La maestra analizó la información anterior y afirmó que 3 corresponde al mayor número de horas diarias que dedican al estudio esos estudiantes. ¿Está usted de acuerdo con esa afirmación?

- A) Sí, porque corresponde al valor máximo.
- B) No, porque el valor máximo es 12.
- C) Sí, porque corresponde a la moda.

Aporte de las TIC

El aporte de las tecnologías de la información y la comunicación (TIC) a la elaboración de ítems se ha evidenciado en el aumento de posibilidades para medir el desempeño en diferentes tareas.

Parshall et al. (2010) presentan una serie de aspectos en que focalizan las innovaciones que pueden tener los ítems de una prueba, en particular, cuando se dispone de medios digitales. Los autores plantean que la presentación en pantalla de una computadora facilita la aparición de formatos novedosos de ítems, así como se cuenta con la posibilidad de incluir nuevos elementos para responder y que brindan mayor complejidad a un ítem, en el sentido de poder medir capacidades superiores.

En este sentido, es posible dar un mayor realismo a los ítems, sobre todo en ir más allá de una presentación solamente textual de la información transitando hacia la inclusión de audios, videos y animaciones.

En un futuro, se espera que la Prueba Nacional Estandarizada se enriquezca con el aporte de las TIC, en particular, al utilizar formatos interactivos.

Diseño Universal de Evaluación

Este es un enfoque que, en pruebas de gran escala, permite una flexibilización en el diseño y desarrollo que permita la participación del mayor número posible de personas estudiantes, así como para hacer inferencias válidas sobre su desempeño en este tipo de evaluaciones (Thompson et al., 2002).

Thompson y Thurlow (2002), así como Thompson et al. (2002) y Hanna (2005) han propuesto una serie de principios para el diseño universal de evaluaciones, los cuales se describen a continuación:

- **Evaluación inclusiva de la población:** Desde el diseño de la prueba, se debe tener como norte la inclusión de toda la población estudiantil, tomando en consideración el amplio rango de condiciones presentes en dicha población.
- **Definición precisa de lo que se va a medir:** Se debe establecer con precisión lo que se va a medir en la prueba, de modo que se eliminen todas las barreras cognitivas, sensoriales, emocionales y físicas irrelevantes.

-
- **Accesibilidad y eliminación del sesgo en los ítems:** Los principios que se refieren a la accesibilidad deben incorporarse al diseño de ítems desde el inicio del proceso y elaborar revisiones que garanticen que se mantenga la calidad en los ítems, eliminando toda forma de sesgo (ventajas o desventajas en la presentación o el contenido del ítem que puedan invalidar su aporte a la puntuación de una prueba).
 - **Adaptaciones:** El diseño de los ítems debe permitir las adaptaciones que se requieran a fin de reducir las amenazas contra las evidencias de validez y comparabilidad de puntuaciones. Por ejemplo, prueba transcrita a sistema Braille.
 - **Instrucciones y procedimientos sencillos, claros e intuitivos:** Todas las instrucciones y los procedimientos son sencillos, claros y se presentan en un lenguaje comprensible.
 - **Máxima legibilidad y comprensibilidad:** Se debe procurar, en la elaboración de ítems, el uso de un lenguaje sencillo (por ejemplo, organización del texto, reducción de la longitud de frases y cantidad de palabras consideradas como difíciles) para que el texto sea legible y comprensible. Este mismo principio aplica para figuras y tablas, así como formatos de respuesta presentes en un ítem.

Para la Prueba Nacional Estandarizada se han seguido los principios del diseño universal de evaluaciones, como una forma de garantizar mayor accesibilidad de la población estudiantil a esta prueba.

Con la colaboración del equipo de personas asesoras del Departamento de Apoyos Educativos para el estudiantado con discapacidad de la Dirección de Desarrollo Curricular, se elaboró el documento “Tabla de codificación de apoyos educativos” para la Prueba Nacional Estandarizada, tomando en consideración tanto la aplicación digital como la aplicación en formato físico.

A continuación, se desglosan los aspectos técnicos de dicho documento, sin explicitar la codificación que tiene un propósito logístico asociado principalmente a la matrícula de las personas estudiantes para la aplicación de la Prueba Nacional Estandarizada.

Organización de la población estudiantil de acuerdo con su condición asociada o no a discapacidad

Se presenta la descripción de las diversas condiciones de la población estudiantil, tal y como se detalla a continuación:

1. **Dificultades específicas del aprendizaje:** (las personas estudiantes con dificultades específicas en la lectura, la expresión escrita, el cálculo y en la adquisición y uso del lenguaje (hablado o escrito); incluye también a personas estudiantes con dislexia,

disortografía, disgrafía, TDA-H, discalculia, trastornos específicos del lenguaje que inciden en la comprensión lectora y la producción escrita, de conformidad con el historial educativo.

2. **Ceguera:** la persona estudiante que tiene pérdida total de la visión.
3. **Baja visión:** el término se refiere a la disminución de la capacidad visual de una persona estudiante, a pesar de utilizar anteojos, tratamiento médico o haberse sometido a una cirugía. La persona está inscrita en el grupo de personas estudiantes con discapacidad visual.
4. **Sordera:** persona estudiante con sordera cuando la Lesco sea su lengua materna, su lengua natural o su lengua de mayor competencia y es la que requiere para el acceso y la comprensión de la comunicación.
5. **Pérdida auditiva:** persona estudiante con pérdida total de la audición en ambos oídos; comprende pérdidas neurosensoriales severas y profundas o anacusias, según audiometría o epicrisis, comprende el uso de prótesis auditivas o audífono y el implante coclear.
6. **Sordoceguera:** discapacidad sensorial, que involucra una disminución significativa de la visión y de la audición, de manera conjunta. Por ejemplo: una persona ciega con hipoacusia, o con baja visión y sordera, o con baja visión e hipoacusia.
7. **Trastorno del Espectro Autista (TEA):** trastorno del neurodesarrollo que comprende una amplia gama de características de funcionamiento diverso, tanto en la comunicación e interacción social, como en comportamientos, pensamientos e intereses rígidos, repetitivos y estereotipados.
Nota: A partir de lo establecido en el DSM-V y la clasificación de Salud CIE 11, otras categorías diagnósticas como el síndrome de Asperger se incluyen dentro del Trastorno del espectro autista.
8. **Discapacidad motora:** limitaciones permanentes del sistema neuromuscular (posturales, de desplazamiento, coordinación de movimientos, expresión oral), debidas a un deficiente funcionamiento en el sistema óseo-articular, muscular y/o nervioso, y que limitan la capacidad funcional en grados muy variables. Las más habituales son la parálisis cerebral, la espina bífida y las distrofias musculares. Comprende también Hemiplejía, Amputaciones, entre otras.
9. **Discapacidad múltiple (Multidiscapacidad o Retos Múltiples):** cuando se presenta más de un tipo de discapacidad en una persona y requieren apoyos extensos y generalizados en las áreas de comunicación, movilidad, vida diaria, socialización y aprendizaje en general.

-
10. **Condiciones asociadas a discapacidad psicosocial** (Esquizofrenia, Bipolaridad, Depresión, Trastorno Obsesivo Compulsivo, Trastorno Negativista Desafiante): condiciones que se caracterizan por alteraciones psíquicas que inciden en el estado de ánimo, las emociones, el pensamiento y el comportamiento.

 11. **Otro tipo de condición:** no es discapacidad, pero requiere de apoyos educativos (Trastornos metabólicos, Trastornos cromosómicos, Enfermedad Neurodegenerativa, Epidermólisis Bullosa (piel de mariposa), Trastorno del habla, Trastornos de la voz, personas oxígeno dependientes, otras situaciones derivadas de una condición de enfermedad crónica (diabetes, epilepsia, fibrosis quística, osteogénesis imperfecta, síndrome nefrótico y trastornos colorrectales), de conformidad con el historial educativo.

Organización de la población estudiantil con apoyos educativos por recintos de aplicación de acuerdo con la condición asociada o no a discapacidad

En el caso de la aplicación digital, se entiende como “recinto aparte” el laboratorio de Informática Educativa o espacio físico adecuado que cuente y cumpla con los recursos tecnológicos necesarios, incluyendo con las ayudas técnicas o los productos de apoyo que garanticen la accesibilidad.

Los productos de apoyo se entienden como cualquier recurso (incluyendo dispositivos, equipo, instrumentos, tecnología y software), fabricado especialmente o disponibles en el mercado, según las normas ISO 9999, utilizados por las personas con discapacidad ya sea de forma independiente o con apoyo personal y destinados a facilitar su participación.

Condición de la persona estudiante	Tipo de recinto	Descripción del recinto
Dificultades específicas de aprendizaje	Laboratorio de Informática Educativa con apoyos educativos	<p>El recinto con un máximo de 10 personas estudiantes, entre los cuales se encuentra un máximo de 05 personas estudiantes, quienes cuentan con apoyos educativos concernientes a las dificultades específicas del aprendizaje.</p> <p>La persona estudiante es atendida por una persona docente de Educación Especial u Orientación Educativa, o bien, con experiencia o formación en el trabajo con esta población estudiantil.</p>

Condición de la persona estudiante	Tipo de recinto	Descripción del recinto
Ceguera	Recinto aparte	<p>Laboratorio de Informática Educativa en atención individualizada con un máximo de 01 persona estudiante ciega (persona que tiene pérdida total de la visión).</p> <p>La persona estudiante es atendida por una persona docente de Educación Especial, específicamente, del área de discapacidad visual, o bien, con formación y/o experiencia en el trabajo con personas estudiantes con discapacidad visual.</p>
Baja visión	Recinto aparte	<p>El recinto atiende a un máximo de 05 personas estudiantes con baja visión.</p> <p>Las personas estudiantes son atendidas por parte de una persona docente de Educación Especial, ya sea específicamente del área de discapacidad visual, o bien, con formación y/o experiencia en el trabajo con personas estudiantes con discapacidad visual.</p>
Sordera	Recinto aparte	<p>El recinto atiende a un máximo de 01 persona estudiante sorda, donde la Lesco es la lengua de mayor competencia (lengua materna o lengua nativa, es decir, aquella que la persona aprende en su infancia y normalmente constituye su instrumento natural de pensamiento y comunicación) y, por tanto, requiere de la Lesco para complementar la comprensión de la prueba. La persona estudiante es atendida por parte de una persona docente de Educación Especial del área de audición y lenguaje, o bien, una persona docente de cualquier otra especialidad y que presente dominio de la Lesco.</p>
Pérdida auditiva	Recinto aparte	<p>El recinto atiende en atención individualizada, a un máximo de 01 personas estudiantes con pérdida auditiva o sordera, donde el Español es la lengua de mayor competencia, pero puede requerir o no, del lenguaje Lesco para complementar la comprensión de la prueba.</p> <p>Las personas estudiantes son atendidas por parte de una persona docente de Educación Especial del área de audición y lenguaje, o bien, una persona docente de cualquier otra especialidad y que presente dominio de la Lesco.</p>
Sordoceguera	Recinto aparte	<p>El recinto con atención individualizada a un máximo de 01 persona estudiante con sordoceguera. La persona estudiante es atendida por parte de una persona docente de Educación Especial con formación y/o experiencia en el trabajo con atender a esta población.</p>

Condición de la persona estudiante	Tipo de recinto	Descripción del recinto
Trastorno del Espectro Autista (TEA)	Recinto aparte	<p>El recinto con atención individualizada, a un máximo de 01 persona estudiante con TEA.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial del área de Problemas Emocionales y de Conducta, o bien, con formación generalista</p>
Discapacidad motora	Recinto aparte	<p>El recinto con atención individualizada, a un máximo de 01 persona estudiante con discapacidad motora.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial generalista, con formación y/o experiencia en el trabajo con esta población.</p>
Discapacidad múltiple	Recinto aparte	<p>El recinto con atención individualizada, a un máximo de 01 persona estudiante con discapacidad múltiple.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial del área de discapacidad múltiple, o bien, con formación y/o experiencia en el trabajo con esta población.</p>
Condiciones asociadas a discapacidad psicosocial	Recinto aparte	<p>El recinto con atención individualizada, a un máximo de 01 persona estudiante con condiciones asociadas a discapacidad psicosocial.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial con formación generalista.</p>
Otro tipo de condición	Recinto aparte	<p>El recinto con atención individualizada, a un máximo de 01 persona s estudiante con otro tipo de condición que no es discapacidad, pero requiere de apoyos educativos.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial.</p>

En la **aplicación en formato físico**, la organización por recintos de aplicación es:

Condición de la persona estudiante	Tipo de recinto	Descripción del recinto
Dificultades específicas de aprendizaje	Aula ordinaria con apoyos educativos que garanticen la accesibilidad	<p>Aula ordinaria con un máximo de 10 personas estudiantes: entre los cuales se encuentra un máximo de 05 personas estudiantes, quienes cuentan con apoyos educativos concernientes a las dificultades específicas del aprendizaje.</p> <p>En el componente de Lenguas Extranjeras, El recinto con un máximo de 01 persona estudiante, que cuenta con apoyos educativos concernientes a las dificultades específicas del aprendizaje.</p> <p>La persona estudiante es atendida por una persona docente de Educación Especial u Orientación Educativa, o bien, con experiencia o formación en el trabajo con esta población estudiantil.</p>
Ceguera	Recinto aparte que garantice la accesibilidad	<p>Atención individualizada con un máximo de 01 persona estudiante ciega (persona estudiante que tiene pérdida total de la visión y para efecto de las pruebas, requiere de estas en Braille).</p> <p>La persona estudiante es atendida por una persona docente de Educación Especial, específicamente del área de discapacidad visual, o bien, con formación y/o experiencia en el trabajo con personas estudiantes con discapacidad visual y que domina la escritura Braille.</p>
Ceguera (casos de discapacidad visual adquirida recientemente)	Recinto aparte que garantice la accesibilidad	<p>Atención individualizada con un máximo de 01 persona estudiante con discapacidad visual adquirida recientemente (persona estudiante ciega que, al adquirir la discapacidad visual, no logró desarrollar un proceso de lectoescritura y que no utiliza la escritura Braille y, por ende, requiere de apoyo para la lectura de la prueba).</p> <p>La persona estudiante es atendida por una persona docente de Educación Especial, específicamente del área de discapacidad visual, o bien, con formación y/o experiencia en el trabajo con personas estudiantes con discapacidad visual.</p>
Baja visión	Recinto aparte que garantice la accesibilidad	<p>Atención a un máximo de 05 personas estudiantes con baja visión.</p> <p>Las personas estudiantes son atendidas por parte de una persona docente de Educación Especial, ya sea específicamente del área de discapacidad visual, o bien, con formación y/o experiencia en el trabajo con personas estudiantes con discapacidad visual.</p>

Condición de la persona estudiante	Tipo de recinto	Descripción del recinto
		<p>En el componente de Lenguas Extranjeras, el recinto atiende a un máximo de 01 persona estudiante con baja visión.</p> <p>Esta persona es atendida por parte de una persona docente de Educación Especial, ya sea específicamente del área de discapacidad visual, o bien, con formación y/o experiencia en el trabajo con personas estudiantes con discapacidad visual.</p>
Sordera	Recinto aparte que garantice la accesibilidad	<p>El recinto atiende a un máximo de 01 persona estudiante sorda, donde la Lesco es la lengua de mayor competencia y, por tanto, requiere de la Lesco para complementar la comprensión de la prueba.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial del área de audición y lenguaje, o bien, una persona docente de cualquier otra especialidad y que presente dominio de la Lesco.</p>
Pérdida auditiva	Recinto aparte que garantice la accesibilidad	<p>El recinto atiende en atención individualizada, a un máximo de 01 personas estudiantes con pérdida auditiva o sordera, donde el Español es la lengua de mayor competencia, pero puede requerir o no, del lenguaje Lesco para complementar la comprensión de la prueba.</p> <p>Las personas estudiantes son atendidas por parte de una persona docente de Educación Especial del área de audición y lenguaje, o bien, una persona docente de cualquier otra especialidad y que presente dominio de la Lesco.</p>
Sordoceguera	Recinto aparte que garantice la accesibilidad	<p>El recinto con atención individualizada a un máximo de 01 persona estudiante con sordoceguera. La persona estudiante es atendida por parte de una persona docente de Educación Especial con formación y/o experiencia en el trabajo con atender a esta población.</p>
Trastorno del Espectro Autista (TEA)	Recinto aparte que garantice la accesibilidad	<p>El recinto con atención individualizada, a un máximo de 01 persona estudiante con TEA.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial del área de Problemas Emocionales y de Conducta, o bien, con formación generalista</p>
Discapacidad motora	Recinto aparte que garantice la accesibilidad	<p>El recinto con atención individualizada, a un máximo de 01 persona estudiante con discapacidad motora.</p> <p>La persona estudiante es atendida por parte de una persona docente de Educación Especial generalista, con formación y/o experiencia en el trabajo con esta población.</p>

Condición de la persona estudiante	Tipo de recinto	Descripción del recinto
Discapacidad múltiple	Recinto aparte que garantice la accesibilidad	El recinto con atención individualizada, a un máximo de 01 persona estudiante con discapacidad múltiple. La persona estudiante es atendida por parte de una persona docente de Educación Especial del área de discapacidad múltiple, o bien, con formación y/o experiencia en el trabajo con esta población.
Condiciones asociadas a discapacidad psicosocial	Recinto aparte que garantice la accesibilidad	El recinto con atención individualizada, a un máximo de 01 persona estudiante con condiciones asociadas a discapacidad psicosocial. La persona estudiante es atendida por parte de una persona docente de Educación Especial con formación generalista.
Otro tipo de condición	Recinto aparte que garantice la accesibilidad	El recinto con atención individualizada, a un máximo de 01 persona estudiante con otro tipo de condición que no es discapacidad, pero requiere de apoyos educativos. La persona estudiante es atendida por parte de una persona docente de Educación Especial.

Descripción de los apoyos educativos

Se advierte que, en todos los casos -sin excepción-, se necesita aportar la evidencia en el historial educativo de la persona estudiante que demuestre su condición y los apoyos educativos que requiere.

En la **aplicación digital**, tales apoyos consisten en:

- a. Hora adicional al tiempo establecido para la realización de la prueba (180 minutos)
- b. Prueba específica: Es una versión de la Prueba Nacional Estandarizada en la que se modifica el 20% de los ítems que sean considerados de mayor complejidad por criterio de personas expertas.
- c. Apoyo físico a la persona estudiante que por su condición requiera de este para consignar en la plataforma la respuesta elegida.

-
- d. Permitir el uso del diccionario en lengua española no enciclopédico y/o diccionario de sinónimos y antónimos para la población estudiantil sorda.
 - e. Inclusión en la gestión de la aplicación de la Prueba Nacional Estandarizada en la plataforma digital de la interpretación en Lesco de las preguntas, por medio de diversos recursos digitales (avatar, hipervínculos, cápsulas de video, otros).
 - f. Persona tutora docente de I y II Ciclos (en primaria) o persona tutora docente de Educación Diversificada (en secundaria) para la Prueba Nacional Estandarizada.
 - g. Uso de hardware (computadora portátil, tableta), opciones de accesibilidad del equipo tecnológico (contraste, ampliación de letra, zoom), software (lector de pantalla, TOBI, comunicador) y otros dispositivos (audífonos) utilizados en el proceso educativo para la implantación y ejecución de la prueba en formato digital.
 - h. Lectura oral de la prueba (apoyo personal o uso de dispositivo tecnológico) a personas estudiantes que, por su condición de discapacidad debidamente comprobada médica y pedagógicamente, así lo requieren.
 - i. Permitir períodos de descanso dentro o fuera del recinto de aplicación de las pruebas, con la supervisión respectiva, a personas estudiantes que por su condición física y socioemocional así lo requieren.
 - j. Para la autorización de este apoyo educativo se requiere contar con la evidencia de su aplicación en el proceso educativo.
 - k. Persona tutora en lengua nativa indígena (si se suministra por parte de la Dirección Regional de Educación).
 - l. Uso de lámpara y calculadora parlante para la persona estudiante con ceguera o baja visión.
 - m. Ubicar a la persona estudiante en un recinto que garantice la accesibilidad.
 - n. Formato digital que permita realizar cambios y ofrecer diferentes opciones de tipo, tamaño de letra (fuente) y contraste.

En la **aplicación en formato físico**, los apoyos consisten en:

- a. Hora adicional al tiempo establecido para la realización de la prueba (180 minutos)
- b. Prueba específica: Es una versión de la Prueba Nacional Estandarizada en la que se modifica el 20% de los ítems que sean considerados de mayor complejidad por criterio de personas expertas.
- c. Hora adicional al tiempo establecido para la realización de la prueba (ciento ochenta minutos).
- d. Apoyo físico a la persona estudiante que por su condición requiera de este para consignar en la hoja de respuesta la opción seleccionada.
- e. Permitir el uso del diccionario en lengua española no enciclopédico y/o diccionario de sinónimos y antónimos para la población estudiantil sorda.
- f. Persona tutora docente de Educación Especial en Audición y Lenguaje en la Prueba Nacional Estandarizada para el II Ciclo de la EGB y Educación Diversificada.
- g. Persona tutora docente de I y II Ciclos (en primaria) o persona tutora docente de Educación Diversificada (en secundaria) para la Prueba Nacional Estandarizada.
- h. Prueba transcrita a sistema Braille.
- i. Lectura oral de la prueba (apoyo personal o uso de dispositivo tecnológico) a personas estudiantes que, por su condición de discapacidad debidamente comprobada médica y pedagógicamente, así lo requieren.
- j. Permitir períodos de descanso dentro o fuera del recinto de aplicación de las pruebas, con la supervisión respectiva, a personas estudiantes que por su condición física y socioemocional así lo requieren.
- k. Persona tutora en lengua nativa indígena (si se suministra por parte de la Dirección Regional de Educación).
- l. Uso de lámpara, lupa, regleta, punzón, pizarra, máquina Perkins, calculadora parlante para personas estudiantes con ceguera o baja visión.
- m. Ubicar a la persona estudiante en un recinto que garantice la accesibilidad.
- n. Prueba Nacional Estandarizada en formato físico con diferentes opciones de tipo, tamaño de letra (fuente) y contraste.

Referencias bibliográficas

- Haladyna, T.M., & Rodríguez, M.C. (2013). *Developing and Validating Test Items*. Routledge.
- Haladyna, T.M., Downing, S.M., & Rodríguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334. https://doi.org/10.1207/S15324818AME1503_5
- Hanna, E. (2005). *Inclusive Design for Maximum Accesibility: A Practical Approach to Universal Design*. Pearson Educational Measurement. http://images.pearsonassessments.com/images/tmrs/tmrs_rg/RR_05_04.pdf
- Herrera Ortiz, M., Benavides Posadas, D. &, Monroy Cazorla, L. (2009). Establecimiento de estándares en un examen criterial.
- Ministerio de Educación Pública de Costa Rica. (2017). *Política Educativa "La persona: Centro del proceso educativo y sujeto transformador de la sociedad"*.
- Ministerio de Educación Pública de Costa Rica. (2016). *Política Curricular "Educar para una nueva ciudadanía"*.
- Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered design: Layers, Structures, and Terminology*. SRI International.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A Brief Introduction to EvidenceCentered Design*. Educational Testing Service.
- Mislevy, R. J., Haertel, G., Riconscente, M., Wise Rutstein, D., & Ziker, C. (2017). *Assessing Model-Based Reasoning using Evidence-Centered Design: A Suite of Research-Based Design Patterns*. Springer.
- Muñiz, J. (2018). *Introducción a la Psicometría: Teoría Clásica y TRI*. Ediciones Pirámide.
- Parshall, C.G., Harmes, J.C., Davey, T., & Pashley, P.J. (2009). Innovative Items for Computerized Testing. En W. van der Linden, & C. Glas (eds). *Elements of Adaptive Testing. Statistics for Social and Behavioral Sciences* (pp. 215-230). Springer.
- Rodríguez, M.C. (2005). Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Measurement: Issues and Practice*, 24(2), 3-13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>
- Rodríguez Frías, M. B., y Flotts de los Hoyos, M. P. (2019). *Definición del referente de la evaluación y desarrollo del marco de especificaciones*. Cuadernillo técnico de evaluación educativa 3. Centro de Medición MIDE UC e Instituto Nacional para la Evaluación de la Educación INEE.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments (Synthesis Report 44)*. University of Minnesota, National Center on Educational Outcomes.

<http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>

Thompson, S., & Thurlow, M. (2002). *Universally designed assessments: Better tests for everyone! (Policy Directions No. 14)*. University of Minnesota, National Center on Educational Outcomes.

<http://education.umn.edu/NCEO/OnlinePubs/Policy14.html>

Zieky, M. J. (2014). An introduction to the use of evidence-centered design in test development. *Psicología Educativa*, 20(2), 79-87, <http://dx.doi.org/10.1016/j.pse.2014.11.003>

Nota: Secciones de este documento han sido amablemente cedidas por el autor cuya referencia es:

Artavia-Medrano, Á. (2014). *Evaluación cognitiva diagnóstica en Matemática: modelo elaborado con el método Rule Space para estudiantes costarricenses de undécimo año*. [Tesis doctoral]. Universidad de Costa Rica.

Autoridades ministeriales

Anna Katharina Müller Castro
Ministra de Educación Pública

Karla Salguero Moya
Viceministra Académico

Sofía Ramírez González
Viceministra Administrativa

José Leonardo Sánchez Hernández
Viceministro de Planificación y Coordinación Regional

Álvaro Artavia Medrano
Director de Gestión y Evaluación de la Calidad

Ana Carvajal Granados
Subdirectora de Gestión y Evaluación de la Calidad

María Elena Parra Arias
Jefa Departamento de Evaluación Académica y Certificación

San José, Costa Rica
Febrero 2024

Equipo técnico

Ciencias

Ramón Montoya Jiménez
Alcira Gámez García
Montserrat Chacón Gómez

Estudios Sociales

Ana Balbina Flores Cornejo
Arnaldo Rodríguez Espinoza

Español

Wilfredo Acevedo Mojica
Ana Jenny Godínez Porras
Álvaro Jiménez Montes
Karla Rubié Morales
Vanessa Zárate Montero

Educación Cívica

Noelia Alvarado González
José Joaquín Rodríguez Chaves

Matemáticas

Elizabeth Figueroa Fallas
Rafael González Palacios
Omar Guzmán Alvarado
Gerardo Murillo Vega

Apoyos educativos

Luis Carlos Rodríguez León



Febrero 2024