



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Propuesta metodológica para la caracterización del Razonamiento Inductivo en la Primera Infancia desde el enfoque de la Educación STEM

Tesis presentada como requisito parcial para optar al título de:
Doctora en Psicología

Presentado Por: **Isabel Astrid Hernández Urrego**

**Director de Tesis: Profesor Hernando Taborda Osorio Ph. D.
Línea De Investigación Construcción De Conocimiento En Contextos Educativos
Doctorado En Psicología Pontificia Universidad Javeriana**

**Codirectora: Profesora Luz Mery González García Ph. D.
Sublínea De Investigación Muestreo Orientado A La Solución De Problemas De Encuestas En
Colombia
Doctorado En Estadística Universidad Nacional De Colombia**

Enero de 2023

Propuesta metodológica para la caracterización del Razonamiento Inductivo en la Primera Infancia desde el enfoque de la Educación STEM

Resumen

El razonamiento inductivo (RI) se distingue de otros procesos cognitivos, por dos características: es fundamental en la generación y transformación del conocimiento, y las inferencias, mediante las cuales se hace explícito, pueden ser total, o parcialmente plausibles. Un importante volumen de evidencia empírica, proveniente de modelos de teorías ingenuas, sugiere que los bebés razonan sobre información estadística, para modificar su comportamiento, elegir y predecir en circunstancias de incertidumbre. Incluso, existe acuerdo sobre las características de experiencias y contenidos que facilitan la inferencia de causalidad, en situaciones probabilísticas cotidianas. Sin embargo, los desarrollos metodológicos, han tardado en llegar a los ámbitos de la formación del pensamiento de los niños habitantes de países pobres.

Desde una perspectiva integradora, se retomaron principios de los modelos de cognición infantil que han considerado las particularidades y diferencias individuales del RI normativo, para el estudio de sus modalidades. Así, se orientó la propuesta metodológica en la creación de criterios de medición e indicadores de eficacia de las inferencias inductivas. Mediante éstos, es posible la estimación de los atributos: plausibilidad, precisión, coherencia y relevancia de la inferencia temprana. En consecuencia, los estudios que fundamentan esta disertación, sustentan la utilidad científica de un instrumento de medición de la eficacia en el RI de niños con edades entre 4 y 6 años. Los resultados de la indagación, también aportan nueva evidencia sobre la influencia de algunos hábitos familiares de formación del pensamiento, en el desarrollo conceptual y del RI eficaz de los participantes.

A partir del logro de los objetivos de investigación, se discute la oportunidad de transferir conocimiento sobre desarrollo cognitivo temprano, hacia la formación de habilidades inferenciales implicadas en la comprensión e indagación científicas, desde Preescolar. Se proyecta, que la estrategia metodológica propuesta para la caracterización del RI en la infancia temprana, tendrá impacto en el cumplimiento de los propósitos de la Educación STEM del siglo XXI.

Palabras claves: razonamiento inductivo eficaz, habilidad inferencial temprana, desarrollo conceptual, socialización de la cognición.

Methodological proposal for the characterization of Inductive Reasoning in Early Childhood from the STEM Education approach

Abstract

Inductive reasoning (IR) is distinguished from other cognitive processes by two characteristics: it is fundamental in the generation and transformation of knowledge, and the inferences, through which it is made explicit, can be totally or partially plausible. A substantial body of empirical evidence, from naive model theories, suggests that infants reason about statistical information to modify behavior, make choices, and predict under uncertain circumstances. There is even agreement on the characteristics of experiences and contents that facilitate the inference of causality, in everyday probabilistic situations. However, methodological developments have been slow to reach the areas of thought formation of children living in poor countries.

From an integrative perspective, principles of child cognition models were taken up, which have considered the particularities and individual differences of the normative IR, for the study of its different modalities. Thus, the methodological proposal was oriented towards the creation of measurement criteria and effectiveness indicators of inductive inferences. Through these, it is possible to estimate the attributes: plausibility, precision, coherence and relevance of early inference. Consequently, the studies that support this dissertation support the scientific usefulness of an instrument for measuring efficacy in the IR of children aged between 4 and 6 years. The results of the investigation also provide new evidence on the influence of some family habits of thought formation, in the conceptual development and the effective IR of the participants.

From the achievement of the research objectives, the opportunity to transfer knowledge, on early cognitive development, towards the formation of inferential skills involved in scientific understanding and inquiry, from Preschool, is discussed. It is projected that the methodological strategy proposed for the characterization of IR in early childhood will have an impact on the fulfillment of the purposes of STEM Education in the 21st century.

Keywords: effective inductive reasoning, early inferential ability, conceptual development, socialization of cognition.

Tabla de Contenidos

Resumen.....	1
Abstract.....	2
Tabla de contenido.....	3
Lista de Tablas.....	8
Lista de Figuras.....	9
Introducción.....	10
Capítulo 1: Revisión de Antecedentes.....	16
Características distintivas del RI.....	17
Atributos del RI en la infancia.....	18
RI y transformación del conocimiento de los niños.....	20
Factores que influyen la plausibilidad de la generalización en la infancia.....	20
Conocimiento conceptual de categorías sociales	21
Inferencia probabilística intuitiva	23
Sensibilidad a patrones de propiedades y regularidades de los eventos.....	23
Sensibilidad de los bebés a patrones estadísticos	24
Inferencia de covariaciones y relaciones causales probabilísticas.....	26
Inferencia probabilística intuitiva y aprendizaje estadístico temprano.....	27
El RI eficaz en la infancia.	28
Atributos característicos de la inferencia inductiva eficaz.....	29
Medición de atributos de la eficacia en el RI.....	31
Evaluación de inferencias eficaces en múltiples modalidades....	32
Formación del RI eficaz para el desarrollo de habilidades inferenciales	34
Inferencia eficaz temprana y generación de conocimiento de áreas STEM	35
Desafíos para la formación de habilidades de inferencia en Preescolar.....	37
¿Qué tipo de factores influyen los procesos cognitivos precursores de las habilidades del pensamiento científico?	38
Influencia social en la generación de representaciones conceptuales significativas.	39
Practica del RI a partir de hechos verificables en la cotidianidad.....	40
Prácticas de socialización de la cognición.....	41

Supuestos teóricos que justifican la propuesta metodológica para caracterizar el

RI temprano.....	42
Capítulo 2: Método.....	47
Resumen de las etapas del proceso metodológico.....	47
Participantes.....	47
Estudio 1 primer etapa.....	47
Estudio 1 primer etapa	48
Estudio 2	48
Estudio 3	49
Creación del IMERIN-v1.....	49
Estudio 1: Diseño y prueba de contenidos para un instrumento de medición de atributos del RI	49
Participantes	52
Diseño.....	53
Procedimiento.....	53
Instrumentos y recursos aplicación de prueba IMERIN-v1.....	53
Resultados de la aplicación de prueba.....	54
Puntuación de la subescala generalización.....	54
Evaluación de desempeño en las actividades.....	54
Estudio 1 -etapa 2. Diseño de nuevos contenidos del IMERIN P.....	59
Participantes.....	59
Instrumentos y Procedimientos.....	59
Etapa 2a.....	61
Etapa 2b.....	61
Etapa 2c.....	61
Conclusiones sobre la confiabilidad de los indicadores de eficacia en la inferencia.....	63
Especificación de criterios e indicadores de evaluación de la conclusión verbalizada.....	64
Contenidos elegidos para fase de generalización del IMERIN -P.....	65
Contenidos fase de prueba.....	67
Fases de las actividades de RI.....	67
Versión del IMERIN para juicio de validación de contenidos.....	69
Estudio 2: Indagación de validez y confiabilidad del IMERIN y sus usos propuestos	69
Etapa 2a: Validación de contenidos: Juicio de expertos	71
Participantes.....	72

Recursos e instrumentos.....	72
Procedimiento.....	74
Conceptos de los jueces respecto a la fundamentación teórica del IMERIN.....	74
2b. Aplicación del IMERIN validación de la propuesta de interpretación de puntajes e índice de eficacia del RI.....	75
Participantes.....	75
Diseño.....	76
Instrumento y materiales.....	77
Procedimiento.....	84
2c. Acuerdo interevaluadoras: validez de criterio de las subescalas de medición y validez predictiva del resultado.....	86
Participantes.....	87
Diseño.....	87
Instrumento y recursos.....	87
Procedimiento.....	88
Estudio 3: Diseño y aplicación de prueba del cuestionario sobre prácticas de socialización de la cognición en la infancia temprana.....	88
Etapa 3a.....	89
Contenidos de la encuesta.....	89
Etapa 3b.....	91
Participantes.....	93
Diseño.....	93
Instrumento y materiales.....	94
Cuestionario de prueba	94
Procedimiento.....	94
Evaluación de la viabilidad de la autoaplicación de la encuesta mediante TIC.....	95
Ajustes al instrumento y procedimiento de la encuesta.....	96
Etapa 3c: Aplicación de la encuesta	96
Participantes.....	96
Diseño.....	97
Instrumento.....	98
Encuesta sobre prácticas de socialización de la cognición en la infancia temprana.....	98
Procedimiento.....	98
Capítulo 3: Resultados.....	99
Estudio de validación del IMERIN y su utilidad científica	99

Estudio 2a. Validación de contenidos por juicio de expertos.....	99
Plan de análisis.....	99
Calificaciones de los ítems de medición de la eficacia en el RI.	100
Estudio 2b. Características psicométricas y resultados de la aplicación para validación del IMERIN.....	102
Plan de análisis.	103
Propiedades psicométricas del IMERIN.	112
Niveles de eficacia en el RI y patrones de agrupación en las puntuaciones.....	118
Interpretación de diferencias en los niveles de eficacia entre los participantes.....	122
Estudio 2c. Acuerdo interevaluadoras: validez de criterio de las subescalas de medición y validez predictiva del índice de habilidad inferencial.....	122
Plan de análisis e interpretación.....	122
Desempeño en los dominios STEM y tendencias en el desarrollo de habilidades inferenciales.....	123
Conclusión preliminar.....	132
Perfil característico del RI eficaz de alto nivel.....	132
Estudio 3. Resultados de la aplicación de la encuesta.....	134
Plan de análisis e interpretación de resultados.....	134
Estudio 3b. Prácticas de socialización de la cognición que influyen el puntaje del IMERIN	139
Estudio 3c. Diferencias en la eficacia del RI y en las prácticas de socialización de la cognición asociadas a variables individuales.....	142
Ejemplo de tendencias distintivas en las prácticas de formación de la cognición.....	146
Discusión y conclusiones	147
Hallazgos y proyección de resultados	147
Alcances y limitaciones de los resultados de la investigación.....	151
Conclusiones.....	153
Referencias	156
Anexo A. Imágenes actividades de Familiarización por clasificación IMERINv-1	179
Anexo B. Imágenes de secuencias en las actividades de categorización del IMERIN-v1	180
Anexo C. Imágenes de secuencias de actividades de generalización por categorización del IMERIN-v1.....	181
Anexo D. Ejemplos de actividades de RI en la modalidad intervención, del IMERIN- v1	182

Anexo E. Ejemplos de actividades de RI en la modalidad elección de evidencia y terminación de secuencias del IMERIN- v1.	183
Anexo F. Estudio de redes semánticas	184
Anexo G. Nube de palabras de las Redes Semánticas y Red de palabras.....	185
Anexo H. Ejemplo de formato para evaluación de actividad IMERIN-P.....	186
Anexo I. Resultados del estudio de concordancia en el juicio de expertos, con coeficiente W de Kendall	187
Anexo J. Gráfico de la matriz de correlaciones policóricas	188
Anexo K. Ejemplo de Actividad retirada del IMERIN-P	189
Anexo L. Análisis de varianza Puntaje total y cada uno de los ítems prueba Kruskal-Wallis.....	190
Anexo M. Correlaciones Sperman entre puntajes en el ítem y edad	191
Anexo N. Promedios puntaje en Dimensiones	192
Anexo O. Gráficos de puntuaciones promedio en las subescalas del IMERIN	193
Anexo P. Gráficos de puntuaciones promedio en las subescalas del IMERIN	194
Anexo Q. Ítems de la encuesta sobre prácticas de socialización de la cognición	195

Lista de Tablas

Tabla 1. Resumen de etapas del trabajo metodológico.....	46
Tabla 2. Criterios de comparación entre respuesta y principios del IMERIN-v1.....	50
Tabla 3. Especificaciones de procesos involucrados en las tareas del IMERIN-v1.....	52
Tabla 4. Características de la evidencia en dominios del IMERIN-P.....	60
Tabla 5. Matriz de especificaciones para evaluación del desempeño en el IMERIN-P.....	62
Tabla 6. Criterios de evaluación de la coincidencia de la conclusión verbalizada y los indicadores de acierto en el ítem. IMERIN-P.	64
Tabla 7. Composición de la muestra de participantes en la aplicación de validación del IMERIN.....	77
Tabla 8. Matriz A de especificaciones conceptuales de los ítems de la encuesta sobre prácticas de socialización de la cognición	90
Tabla 9. Matriz B de especificaciones para la creación de ítems para la encuesta sobre prácticas de socialización de la cognición.....	91
Tabla 10. Categorías de origen de los ítems de la Encuesta sobre prácticas de socialización de la cognición en la infancia temprana.....	92
Tabla 11. Puntajes totales en las dimensiones de eficacia	93
Tabla 12. Calificación Promedio en los indicadores de actividades del IMERIN.....	101
Tabla 13. Frecuencias y porcentajes calificación ítems dominio Ciencias (N=8).....	113
Tabla 14. Frecuencias y porcentajes calificación ítems dominio Estadística	114
Tabla 15. Frecuencias y porcentajes calificación ítems dominio Tecnología	115
Tabla 16. Medias, distribución y discriminación de los puntajes IMERIN.....	118
Tabla 17. Valores p del estadístico H-prueba K subescalas del IMERIN.....	119
Tabla 18. Medias puntajes en subescalas del IMERIN en subgrupo de edades	121
Tabla 19. Medias de los puntajes en subescalas del IMERIN discriminados por género.....	122
Tabla 20. Prácticas de socialización de la cognición habituales entre los participantes.....	136
Tabla 21. Asociaciones estadísticamente significativas entre prácticas de socialización de la cognición y puntajes del IMERIN.....	140
Tabla 22. Frecuencias de las prácticas de socialización de la cognición habituales y los puntajes promedio en subescalas de eficacia del IMERIN correspondientes	141

Lista de Figuras

Figura 1. Ciclo de la formación de pensamiento transformación del conocimiento STEM.....	36
Figura 2. Actividades de categorización del IMERIN –v1 fase de familiarización.....	55
Figuras 3. Actividades de categorización IMERIN-v1 fase de familiarización.....	56
Figuras 4 y 5. Actividades de categorización por emparejamiento	57
Figuras 6 y 7. Imágenes de las secuencias generalización por categorización, IMERIN -v1.....	58
Figuras 8 y 9. Ejemplos de los ítems de generalización del IMERIN- P	66
Figuras 10. Ejemplo de una actividad del IMERIN-P	68
Figuras 11. Ejemplo de una actividad del IMERIN-P.....	70
Figuras 12 a 17. Tres (de 6) actividades del IMERIN, versión validada por jueces.....	83
Figuras 18 a 26. Dos actividades del IMERIN con explicación de calificaciones.	104
Figura 27. Gráfico de Coeficientes de correlación policórica interdominios.....	117
Figura 28. Gráfico de la estructura de componentes principales del IMERIN.	118
Figura 29. Gráfico de niveles de puntuación en las dimensiones de eficacia en el RI	120
Figura 30. Estructura de clasificación de patrones distintivos del desempeño del IMERIN, o “Árbol de decisión”...	124
Figura 31. Clasificación de desempeño en el Arbol de decisión ítems significativos.....	126
Figura 32. Árbol de decisión imágenes en los ítems distintivos del desempeño IMERIN.....	127
Figuras 33. Actividad de explicación Ítem de generalización, del dominio Tecnología, con interpretación de calificaciones de 6 participantes.	128
Figuras 34. Actividad de explicación Ítem de precisión, del dominio Tecnología, TP1, con interpretación de calificaciones de 6 participantes.....	129
Figuras 35. Actividad de explicación Ítem de coherencia, del dominio Tecnología, TC1, con interpretación de calificaciones de 6 participantes.....	130
Figuras 36. Ítem de relevancia de la actividad explicación, en el dominio Tecnología, TR1, con interpretación de calificaciones de 6 participantes.....	131
Figuras 37. Arbol de decisión: prácticas de socialización y variables individuales asociadas...	143
Figuras 38. Arbol de decisión: prácticas de socialización y puntaje promedio del IMERIN.....	145

Introducción

Actualmente se debate si habilidades inferenciales fundamentales para la comprensión en dominios específicos, como la Ciencia, se desarrollan temprano en la infancia, o si su desarrollo es más tardío y se prolonga durante los años de escolarización inicial, incluso hasta la adolescencia. Se justifica abordar esta problemática, porque la formación de estas habilidades de pensamiento, fomenta el desarrollo y comprensión del conocimiento de dominios STEM (Science, Technology, Engineering and Mathematics), incluso desde Preescolar (Epstein y Fisher, 2017; Friedmab, 2005; Herro y Quigley, 2017; Sing, 2021).

Esta cuestión se asocia con inconsistencias entre algunos hallazgos en la indagación del desarrollo cognitivo temprano, en las últimas décadas del siglo XX, y de estudios de caracterización de los mecanismos inductivos, de años recientes (Bonawitz et al., 2019; Carey, 1985; Gopnik y Meltzoff, 1997; Wellman y Gelman, 1998). Es probable que, la falta de unidad en la conceptualización de los procesos subyacentes al desarrollo del pensamiento científico, desde la infancia, también se relacione con discrepancias frente a la naturaleza del vínculo entre el cambio conceptual y el *razonamiento inductivo*, RI en adelante (Carey, 2009; Gopnik y Tenenbaum, 2007; Hayes, McKinnon y Sweller, 2008).

Perspectivas actuales del RI, lo definen como el proceso de coordinar inferencias que se generan intencionalmente, siguiendo normas para favorecer la plausibilidad o probabilidad de la conclusión (Johnson y Ahn, 2017; Ricco, 2015). Las conclusiones derivadas del RI explícito se describen como inferencias, completa o parcialmente plausibles (Colberg, Nester y Trattner, 1985; Nisbett y Ross, 1988). De manera que, algunas de las inferencias inductivas son mejores que otras, o más razonables (Moshman, 2004; Markovits y Barrouillet, 2004).

En ese marco, la definición operacional del RI debe considerar comportamientos tales como: *representar*; que significa hacer modelos mentales del significado de las cualidades distintivas de las categorías y de las relaciones entre sus propiedades (Johnson-Laird, 1980, 1994); e *inferir nuevo conocimiento* o información, a partir del contenido objeto de razonamiento (Holland et al., 1983; Medin et al., 2003). De manera que, la característica distintiva fundamental del RI

explícito es que se expresa mediante inferencias, o conclusiones que difieren en su grado de justificación. Estas, a su vez, son evidencia del isomorfismo entre los modelos mentales del razonador y los principios que definen las relaciones razonadas (Holland et al, 1989; Johnson-Laird, 1994).

Diferentes modelos del razonamiento enfatizan que la disponibilidad de representaciones semánticas, coherentes, es un factor crítico de la plausibilidad de las hipótesis y conclusiones. Especialmente, si éstas son logradas en condiciones de incertidumbre (Köksal-Tuncer y Sodian, 2018; Leighton y Sternberg, 2004; Xu y Kushnir, 2012). En condiciones de certeza limitada, la plausibilidad de la inferencia puede ser consecuencia de la capacidad asociar el conocimiento de categorías y la representación de diversas relaciones entre éstas (Koloswky, 1996). Así mismo, la probabilidad de las conclusiones podría depender de la coincidencia entre la evidencia y los supuestos, sobre el caso razonado, que integra el razonador al proceso (Stenning y Monaghan, 2004).

Respecto a los procesos característicos del RI, Leighton y Sterberg (2004), argumentan que durante la inferencia inductiva se razona sobre las representaciones de los eventos, objetos y símbolos, lo cual implica tres formas de procesamiento: a) traducir la entrada sensorial en una representación semántica; b) transformar una representación conceptual en otra más sofisticada; y c) traducir una representación, semántica o conceptual, en una respuesta observable o conducta explícita. Estos procesos de la inferencia inductiva son componentes centrales del aprendizaje humano; puesto que involucran el *uso y transformación del conocimiento disponible*, para hacer conclusiones que tienen probabilidad de integrar nuevos conocimientos (Evans, 2020; Hayes, McKinnon y Sweller, 2008; Moshman, 2004; Nisbett et al., 1983; Rhodes, Gelman y Brickman, 2008).

Las tres formas de procesamiento referidas, ilustran cómo el cambio en el desarrollo de los procesos inferenciales conlleva al cambio en los sistemas conceptuales. Desde esta perspectiva, la traducción de la entrada sensorial, o percepción, en concepto, puede ser entendida como un *cambio semántico* (Waxman y Leddon, 2010). El cambio en las representaciones de información ocurrido durante la inferencia inductiva, es un *cambio conceptual* (Csapo, 1997; Opfer y Siegler, 2004). Así mismo, al *transformar conceptos* en conducta observable, se genera un cambio en el conocimiento, de implícito a explícito (Dienes y Perner 1999). En consecuencia, las inferencias explícitas expresan cualidades distintivas del razonamiento que llevó dichos cambios.

El estudio empírico de los atributos de la inferencia inductiva, en sus diferentes modalidades y en dominios específicos, ha originado teorías y modelos explicativos únicos. En ese sentido, existe acuerdo entre los investigadores que proponen a las teorías intuitivas (ingenuas) como recursos para la transformación de conocimiento de dominios específicos. Así mismo, en la generación de nuevas representaciones semánticas y conceptuales precisas, coherentes y relevantes (Brandone, 2017; Carey, 2009; Heit, 2000; Pillow, 2002; Sutherland y Cimpian, 2017).

Una tendencia en los estudios del desarrollo de habilidades inferenciales, durante las últimas décadas del siglo XX, fue el auge de paradigmas centrados en la generalización inductiva. Esta modalidad del razonamiento se define como la capacidad de inferir patrones distintivos, o reglas de relación entre instancias y propiedades de las categorías, integrándolas a representaciones semánticas y conceptuales generales (Gelman y Davidson, 2013; Rhodes y Brickman, 2008; Schulz et al., 2008; Sutherland y Cimpian, 2017). Dado que la generalización es posible por razonamiento deductivo, resulta impreciso considerar que es un proceso característico de la habilidad inferencial inductiva (Colberg et al., 1982; Colberg et al 1985; Evans, 2020; Gopnik, 2011).

La particularidad que distingue la generalización inductiva es que la conclusión agrega nueva información a las instancias, supuestos o premisas. Especialmente, en situaciones de razonamiento con evidencia indirecta y pocos datos (Evans, 2020; Hayes, 2007; Tenenbaum, 2011). En ese tipo de situaciones, la justificación de la conclusión está dada por su coincidencia con *indicadores de acierto*, o algún tipo de referente de dominio específico con el que resulte pertinente compararla, para verificarla. El conocimiento social y abstracto podrían explicar algunas generalizaciones plausibles, en situaciones de RI cotidiano y en las que no se dispone de conocimientos de dominio específico, por ejemplo (Nisbett et al., 1983)

En lo corrido de este siglo, la Psicología y otras Ciencias Humanas han innovado las metodologías, dando paso a nuevos modelos y paradigmas para la indagación del desarrollo cognitivo en la infancia. En la actualidad, se dispone de evidencia empírica que demuestra, suficientemente que, entre las modalidades de la inferencia inductiva, además de generalizaciones, están incluidas: juicios sobre causas probables, nuevas hipótesis, explicaciones y predicciones basadas en evidencia (Evans, 2020; Goswami, 2010; Hayes et al., 2008; Kim et al., 2012; Legare y Lombrozo, 2014; Lombrozo, 2016).

Respecto a los juicios probabilísticos, el modelo bayesiano de razonamiento explica que los niños hacen inferencias inductivas racionales en el dominio psicológico, de forma consistente con

los principios generales de la inferencia bayesiana. Es decir, se afirma que los niños (incluso menores de 6 años) están en capacidad de expresar creencias en términos de hipótesis que se verifican mediante experiencias de razonamiento basadas en evidencia, o que involucran probabilidades. Desde esta perspectiva del RI, la capacidad de hacer inferencias causales, incluso si los indicadores de causalidad se presentan de forma probabilística, se desarrolla temprano en la infancia (Denison y Xu, 2012; Koslowski y Masnick, 2010; Xu y Griffiths, 2011).

Los argumentos de las líneas bayesianas, sustentan que la inferencia inductiva favorece la revisión de creencias y el cambio conceptual de los razonadores (Lane y Shafto, 2017; Lawson, 2017; Waldmann y Hagmayer, 2006). Este planteamiento ha respaldado la creación de situaciones de RI para el estudio de las habilidades inferenciales de los niños, especialmente en contextos de causalidad, con diversas evidencias y en múltiples dominios (Gopnik y Tenenbaum, 2007; Schulz et al., 2008). Las situaciones de RI, multifacéticas, permiten indagar variabilidad en las estrategias de inferencia y en los conocimientos que usan los niños para hacer conclusiones plausibles, en diferentes modalidades.

Es importante resaltar que, la evaluación de los *atributos de la eficacia* de las inferencias inductivas del niño, en edad preescolar (4 a 6 años), indicará cambio en el desarrollo de los procesos de pensamiento que favorecen el aprendizaje de principios de las investigaciones científicas (OECD, 2009, 2013, 2015). De hecho, como lo explican Evans y Over (2013), se ha determinado que el conocimiento científico actual, así como el conocimiento cotidiano, es probabilístico y tiene origen en inferencias inductivas (amplias). Las cuales, a su vez, expresan conclusiones hipotéticas, que admiten incertidumbre (Wild y Pfannkuch, 1999). Con estos propósitos, algunos estudios empíricos de la cognición infantil indagan el desarrollo, temprano, de la capacidad de usar patrones de información estadística.

Los estudios de atributos del RI eficaz resultaron importantes para la presente disertación porque dicha capacidad, facilita el aprendizaje y generalización de habilidades de inferencia probabilística, subyacentes a la comprensión e indagación científicas (Culbertson, 2016). Con base en estas orientaciones teóricas se abordó la pregunta: ¿cómo se describe el vínculo entre la eficacia de la inferencia inductiva temprana y el desarrollo de habilidades del pensamiento científico?. Para abordar esta cuestión, se adoptó un enfoque normativo para la medición de atributos del RI (Ricco, 2015). Caracterizar procesos inductivos normativos y sus mecanismos subyacentes, aporta conocimiento útil para la creación de tecnologías pedagógicas dirigidas a fomentar el cambio y

desarrollo conceptual. Incluso, este conocimiento orientaría objetivos de modificación de la eficacia del razonamiento con contenidos STEM y de formación temprana del pensamiento científico. Estos procesos han sido planteados por investigadores de la Psicología Educativa (Kuhn, 2010; Mohlhenrich et al., 2018), por divulgadores científicos (OECD, 2009) y por los estándares de Educación del Ministerio de Educación Nacional de Colombia (MEN) como meta del sistema de Educación para el siglo XXI (MEN 2011, 2014).

Se asumieron también, supuestos sobre la influencia cultural en algunas diferencias individuales en el desarrollo del RI eficaz. La influencia de los adultos formadores del pensamiento, puede darse a través de procesos de socialización de la cognición y de la mediación en la búsqueda del conocimiento STEM. Nisbett y Ross (1980) plantean que los sistemas de pensamiento tienen origen en prácticas culturales diversas. De hecho, la ideología de la comunidad influye directamente la plausibilidad de los supuestos metafísicos que usan sus miembros para razonar. Por ejemplo, entre esos supuestos están incluidas las creencias relativas a la diversidad de las muestras y preferencias por enseñar - aprender sobre ciertas categorías y dominios específicos.

Probablemente, en la infancia, las creencias se vinculan con prácticas de formación de la cognición, especialmente en las interacciones del hogar y con la familia. Su indagación es importante, porque algunos antecedentes indican que, a pesar de que este tipo de representación mental podría apoyar generalizaciones plausibles (Lawson y Fisher, 2009), con frecuencia se asocia con fuentes de sesgos (Jacobs y Narloch, 2001) y estereotipos sobre contenidos de dominios específicos (Bascandzhev y Harris, 2016; Bogdan y Meneses, 2019).

En ese orden de ideas, los resultados de la indagación de prácticas tempranas de socialización de la cognición de los niños, podrían sugerir nuevas formas de abordar los interrogantes referidos al aprendizaje y transferencia de representaciones de categorías y conocimiento abstracto, antes de los 6 años (Brown y Campione, 2004; Gopnik y Tenenbaum, 2007; Hayes et al., 2008; Tenenbaum et al., 2011; van Laar et al., 2020). Estas son cuestiones de interés en el ámbito de la Educación inicial, puesto que dichos procesos son bases de las habilidades cognitivas para la comprensión del conocimiento científico y de su función en la cultura (Brenneman y Louro, 2008; UNESCO, 2017; Vega, 2018), entre otros tipos de comprensión social.

La influencia del conocimiento social en las habilidades inferenciales tempranas es un interés investigativo actual en las ciencias cognitivas. Algunas líneas de investigación en Psicología de la Educación sugieren nuevos enfoques, para indagar asociaciones entre la información que reciben

los niños cotidianamente, plausibilidad de sus generalizaciones y justificaciones para sus creencias. Resaltan además, la importancia crítica de esclarecer la función de la mediación del adulto en la disponibilidad de evidencia y de datos relevantes para el RI sobre hechos verificables. Especialmente, si en la inferencia se involucra la confianza que siente el niño hacia el conocimiento de otras personas (Einav y Robinson, 2010; Koenig et al., 2015).

Este panorama general de los antecedentes de investigación del desarrollo cognitivo, sitúa el marco teórico para la caracterización del RI, en tres tipos de actividad inductiva principalmente: generalización inductiva basada en categorías, inferencia inductiva de relaciones causalidad y procesos inferenciales inductivos implicados en el pensamiento científico. Entre los factores que determinaron la elección de las modalidades de inferencia a estudiar, influyó mayoritariamente el hecho de que, en el siglo XXI, han proliferado los reportes de investigación sobre el desarrollo de cada una, sin embargo, pocos contaron con la participación de poblaciones infantiles habitantes de países latinoamericanos.

Aunque se dispone de abundante evidencia sobre el vínculo de la inferencia generalizante y el cambio conceptual, en dominios específicos, ésta es resultado de indagaciones con población de países industrializados. De manera que, para el estudio empírico de la relación entre RI eficaz y habilidades inferenciales de niños residentes de países en vía de desarrollo, es necesario diseñar y validar nuevos instrumentos. Por lo tanto, se justifica proponer nuevas metodologías para la caracterización del RI temprano.

Con el propósito de abordar las posibles respuestas a las cuestiones metodológicas planteadas, se propusieron actividades de RI, que integran la evaluación de la plausibilidad de las generalizaciones y la estimación de la eficacia de inferencias posteriores, observadas en diferentes modalidades. Este enfoque permite indagar la variabilidad intraindividual, e interindividual, en la habilidad inferencial. Dado que, a cada modalidad de la inferencia inductiva corresponde un repertorio de reglas o estrategias, las cuales aplica selectivamente el razonador (Kuhn y Dean, 2004).

En este contexto, investigadores del desarrollo cognitivo y expertos en las disciplinas STEM, plantean que es prioridad centrar las experiencias pedagógicas, formales e informales, en la formación temprana de estrategias de inferencia eficaces (Moshman, 2004; Ricco, 2015). Puesto que, las situaciones de evaluación de las estrategias y reglas de inferencia, pueden generar evidencia de la configuración de patrones de información teórica, verificable (Kuhn y Dean, 2004)

y de creencias tanto relevantes como sesgadas (Kuhn, 2006; Evans y Over, 2013). La presente propuesta metodológica corresponde con el propósito de apoyar la creación e innovación de dichas experiencias antes y durante el Preescolar.

El objetivo general de la investigación fue proponer un método para la caracterización del RI de niños colombianos, con edades entre 4 y 6 años, desde el marco teleológico de la Educación STEM. Los objetivos específicos son: 1) medir la eficacia en la inferencia inductiva de los niños a partir de sus conclusiones explícitas, usando contenidos de los dominios STEM; 2) avanzar en la creación de estrategias de evaluación del RI en función del desarrollo y cambio conceptual en la primera infancia; 3) describir implicaciones de la influencia de las prácticas familiares de socialización de la cognición en la toma de decisiones pedagógicas en el Preescolar.

En el capítulo 1 se presenta una síntesis de los antecedentes conceptuales y pragmáticos de sustento a los casos teóricos que originaron las hipótesis y preguntas de investigación. En el capítulo 2 se describen los 3 estudios llevados a cabo para el diseño y validación de los instrumentos de estimación de atributos del RI eficaz, y para indagar algunas prácticas familiares de formación de la cognición. Los resultados se presentan y analizan en el capítulo 4, así como las conclusiones, discusión de alcances y proyecciones para el uso del conocimiento aportado en esta disertación.

Capítulo 1: Revisión de Antecedentes

Con el fin de lograr una aproximación objetiva al estado del arte de la investigación empírica del RI en la infancia se siguió la metodología llamada Revisión Sistemática de Literatura, descrita por Petticrew y Roberts (2006). La búsqueda inicial resultó en la selección de 336 textos, entre tesis, libros y artículos arbitrados. La mayoría de las publicaciones (más del 97%) están escritas en Inglés. Se depuró la base para la sistematización, priorizando estudios empíricos del desarrollo cognitivo en la infancia temprana, realizados durante las últimas tres décadas. Finalmente se analizó cada elemento de la muestra, delimitando referentes conceptuales y metodológicos; actualizados, coherentes y verificables empíricamente, para la caracterización del RI en la infancia.

Características distintivas del RI

El RI es la adaptación singular del pensamiento al contexto (Nisbett et al., 1983; Nisbett et al., 2001). Se define como el proceso mediante el cual se generan inferencias usando conocimientos

relativos a la evidencia y a los principios que explican las relaciones en la situación razonada (Leighton, 2004). Dado que el RI es ubicuo a la actividad cognitiva y aplica el conocimiento representado (Shipley y Shepperson, 2006; Stenning y Monaghan, 2004), se considera fundamental para abordar soluciones a diferentes problemas, especialmente, a los cotidianos. En ese orden de ideas, las inferencias inductivas son el tipo de conclusión que genera el razonamiento en circunstancias novedosas, inesperadas o inciertas (Evans, 2020).

Shafto, Coley y Vitkin (2015) definen la inferencia inductiva como el razonamiento que utiliza el conocimiento conceptual, con el objetivo de evaluar la probabilidad de que una nueva propiedad sea compartida por los diversos elementos de un conjunto de categorías. Dado que la inferencia de nuevo conocimiento es el proceso fundamental del RI, diferentes respuestas explícitas que expresan conclusiones abarcan más características distintivas de este tipo de razonamiento (Markovits y Barrouillet, 2004). Mediante RI también se hacen inferencias predictivas: sobre la probabilidad de relaciones o eventos futuros, de los que se conocen condiciones de ocurrencia, aunque cambien algunas circunstancias (Setoh et al., 2013). Así mismo, se razona sobre la probabilidad de cambios en condiciones pasadas, que explican cambios en la ocurrencia de la relación o evento en el presente. Este es un tipo de RI con fines de diagnóstico.

Algunas definiciones, resaltan la naturaleza probabilística del RI, dado que éste se fundamenta en procesamiento de información sobre la cual solo se tiene algún nivel de certeza. Rehder (2007) explica que el RI es la actividad cognitiva por la cual se generan inferencias que, siendo inciertas, tienen probabilidad de ser generalizadas a una categoría completa de eventos, o elementos. Una modalidad del RI ampliamente estudiada es la inferencia generalizante. Esta se define como la actividad cognitiva que origina generalizaciones de nuevas instancias, desde instancias conocidas (Shafto et al., 2007). Además, se generaliza conocimiento de casos (proposiciones) a otros casos particulares, a través de representaciones más amplias (Stenning y Monaghan, 2004).

En la generalización se pueden diferenciar juicios sobre similitud, que se derivan de inferir semejanza entre los casos o las propiedades de las instancias (o premisas) y juicios referidos a la conclusión. Así mismo, juicios sobre la cobertura, centrados en inferir si las propiedades o instancias de las categorías de la premisa abarcan suficientemente los casos de las categorías de la conclusión (Osherson et al. 1990). La consideración de estos componentes confiere plausibilidad a las inferencias generalizantes.

De acuerdo con la modalidad del RI estudiada se diseñan situaciones y se presentan evidencias que requieren usar algún tipo de representación, o de información específica. De manera que las inferencias sobre cada situación, implican transformar la información de los contenidos, probablemente integrando otros tipos de información. Por lo tanto, el RI explícito se caracteriza por la adición de nuevos supuestos sobre la situación, basados en la valoración de la evidencia (Stenning y Monaghan, 2004), los cuales se expresan mediante conclusiones. Este procesamiento ocurre por la intención de generar nuevas representaciones semánticas o conceptuales sobre la situación inicial.

Otras características distintivas del RI se desprenden de la conceptualización de la inferencia en términos de la transformación, o generación de nuevo conocimiento. Parece existir acuerdo en que este tipo de razonamiento es multidimensional. Por lo tanto, es fundamento de otros procesos cognitivos, por ejemplo, juicios generalizantes sobre categorías (Fisher et al., 2015; Medin, 2003; Osherson, 1990), inferencias en situaciones de causalidad probabilística (Bonawitz et al., 2019) y toma de decisiones (Blok, Osherson y Medin, 2007). Este conjunto de características puede ser traducido a indicadores de desempeño, con el fin de indagar empíricamente los atributos de las conclusiones plausibles y la eficacia de las inferencias.

Atributos del RI en la infancia. Los fundamentos conceptuales para definir y caracterizar el RI en la infancia han cambiado en las últimas décadas. Desde que Susan Carey, en 1985, realizó el primer experimento que usó evidencia diversa para estudiar la inferencia inductiva, con niños menores de 6 años (Carey, 2009; Goswami, 2003, 2010). Actualmente se acepta que, en la infancia temprana, la sensibilidad a los principios de la inducción: similitud y diversidad, pueden ser factores críticos de la plausibilidad de las conclusiones. Así mismo, se ha determinado la importancia de situaciones de razonamiento que incorporen conocimiento conceptual (de categorías y causalidad) en diferentes contextos y con contenidos de dominios específicos (Hayes, 2007; Heit, 2007).

Ricco (2015) y Moshman (2004) definen el razonamiento como un conjunto de inferencias intencionales que se coordinan a partir de principios normativos (para el razonamiento correcto). Así, las conclusiones alcanzan mayor o menor justificación como resultado de considerar dichos principios (o ser sensible a éstos, en el caso de los bebés) (Lawson, 2004). En este marco, se espera (teóricamente) que el RI infantil: a) concuerde con las condiciones contextuales (Markovits, 2004); b) use conocimiento de similitud entre categorías y sus propiedades (Lawson, 2017; Godwin,

Matlen y Fisher 2013); c) considere la diversidad en la evidencia (Shipley, 2006); d) incorpore, coherentemente, conocimiento abstracto (Shafto, 2008); y, e) adicione información relevante, en diferentes dominios (Nguyen et al., 2016).

Estas definiciones integran elementos normativos para el RI explícito, los cuales se refieren a que las conclusiones se diferencian en la probabilidad de ser razonables (Blok et al., 2007; Evans, 2020). De hecho, la plausibilidad de las generalizaciones, precisión, coherencia y relevancia del razonamiento se evalúa con el fin de determinar si las inferencias de los niños se vuelven más eficaces durante el desarrollo (Moshman, 2004; Ricco, 2015). El interés en evaluar la eficacia del RI, corresponde con el acuerdo teórico en que, los niños (desde los 4 años, aproximadamente) están en capacidad de razonar sobre el conocimiento aprendido de la experiencia (conocimiento semántico). Incluso, lo integran a representaciones mentales (conocimiento representado), con el fin de responder, adaptativamente, a los requerimientos del contexto (Sloutsky et al., 2015; Nguyen, et al., 2016; Ricco, 2015).

El RI eficaz permite la revisión de creencias y transformación de representaciones conceptuales, actualizando, ampliando y generando nuevos conocimientos. Puesto que, la conclusión presenta la hipótesis que el razonador cree plausible y es resultado de evaluar evidencias (Evans y Feeney, 2004). El nuevo conocimiento es útil, además, para controlar, o resolver la incertidumbre frente a las relaciones que explican el evento, o resultado objeto de razonamiento (Holyoack y Morrison, 2012; Jhonson-Laird, 1994; Leighon y Sternberg, 2004). Así, la evaluación de la eficacia del razonamiento tiene sentido en situaciones en las que se desconocen reglas de las relaciones, o cuando la evidencia de intervención de diferentes variables en un sistema es indirecta.

RI y transformación del conocimiento de los niños

En la inferencia inductiva se requiere transformar representaciones conceptuales a partir de información relevante, implícita y explícita. Dicha información, está disponible en el contexto cotidiano de los niños, mucho antes de que lleguen al aula (Baillargeon et al., 2010; Baillargeon et al., 2016; Cosmides y Tooby 2013). Este tipo de razonamiento es un recurso cognitivo para la adquisición de nuevos conocimientos, para transformar el aprendido y para facilitar que el conocimiento recordado sea aplicado en nuevos contextos (Csapo, 1997).

Algunas líneas de investigación del desarrollo cognitivo en la infancia, plantean que no hay cambios fundamentales en la naturaleza del RI desde la edad preescolar (Gopnik y Tenenbaum,

2007; Hayes, 2007; Hayes et al., 2008). De manera que, estudian el RI mediante tareas experimentales, con estructuras inferenciales similares, comparando entre sí las conclusiones de niños y adultos (Goswami, 2003, 2010; Sternberg y Gardner, 1983; Shipley y Shepperson, 2006; Rhodes, et al., 2008; Baillargeon, Scott y Bian, 2016; Fizke et al, 2017). También se han creado diversos contenidos, tanto en dominios generales, como específicos, puesto que se acepta ampliamente, que el RI es inherente a toda actividad cognitiva en la infancia temprana.

Diferencias individuales en las estrategias de inferencia, usadas para abordar soluciones a problemas del mundo real, también se explican por diferencias en la disponibilidad de conocimiento semántico y conceptual (Goswami, 2003, 2010; Kuhn, 2005, 2010b; Su et al., 2002). De hecho, el conocimiento y representaciones conceptuales relativas a las interacciones cotidianas entre los agentes de la convivencia pueden ser característicos de la actividad inferencial en la infancia (Heit, 2000; Heyman y Gelman, 2000; Pillow, 2002).

Factores que influyen la plausibilidad de la generalización en la infancia.

Las situaciones de generalización que presentan evidencia de sistemas de relaciones relevantes en un resultado, permiten examinar la transformación de las representaciones intuitivas en conocimiento de variables específicas de un dominio (Booth, 2014; Carey, 2009). En consecuencia, y a partir de la coincidencia entre resultados de estudios de la inferencia generalizante, se plantea que las experiencias de conocimiento de categorías diversas, influyen de forma determinante el cambio y desarrollo conceptual (Quinn, 2010; Waldmann y Hagmayer, 2006). Por lo tanto, generan evidencia relevante para abordar la cuestión del aprendizaje, transferencia y modificación de las habilidades inferenciales.

En reportes de estudios de la capacidad de generalización es común la idea de que expertos y novatos, niños y científicos, razonan inductivamente con *la intención* de inferir el factor causal pertinente (plausible). Es decir, se entiende que el propósito de las conclusiones generalizantes es que den cuenta de la naturaleza ontológica, o del origen de los resultados, las relaciones, o sistemas objeto de RI (Moulines, 2011). En esta línea, Pillow (2002) explica que por RI los niños generan y transforman conocimientos generalizables a partir de la experiencia social. Dichos conocimientos también se logran por inferencia en dominios específicos regulados por principios científicos. Incluso a partir de pocos casos, o con muestras pequeñas de las propiedades y eventos (Reheder,

2007). Gweon, Tenenbaum y Schulz, (2010) hicieron demostraciones probabilísticas para investigar la generalización de propiedades a partir de la observación de eventos de muestreo. Otros estudios que usaron muestreo, con niños de 2 a 5 años de edad (y adultos), confirmaron cambios en generalizaciones asociados al conocimiento específico de muestras diversas, como frecuencias explícitas y propiedades de categorías (Gelman et al., 2015; Rhodes et al., 2008), o por efecto del tamaño (Lawson, 2014) y composición (Shipley y Shepperson, 2006) de las muestras.

La literatura especializada coincide en que, la generalización correcta, implica la comprensión de las propiedades distintivas y de las estructuras de relaciones entre las categorías de diversos dominios. Siendo esta la modalidad de inferencia mayoritariamente indagada en estudios empíricos del razonamiento infantil, es pertinente afirmar que la generalización es la forma de inferencia inductiva fundamental en la infancia. Sin embargo, se han caracterizado otros tipos de juicios explícitos que evidencian la sensibilidad a principios normativos en el RI de los niños (Evans, 2020).

Conocimiento conceptual de categorías sociales. Los bebés transfieren conocimiento social, de eventos regulares, en función del reconocimiento de semejanzas y singularidades en otros sistemas de relaciones, en diferentes dominios (Erikson et al., 2010; McCormack et al., 2016). Los niños también razonan desde la perspectiva longitudinal pasado-futuro (Marcovitch et al, 2015; Walker y Lombrozo, 2014)). Para anticipar y predecir condiciones, o estados emocionales de los agentes involucrados en una situación problemática (Jara-Ettinger, et al., 2015). Es posible que infieran las posibles razones que justifican las acciones de estos agentes (Ruffman et al., 2012).

Las inferencias que los niños logran haciendo uso de conocimiento de los dominios social y psicológico son diversas. Incluyen, por ejemplo: hacer hipótesis sobre estados mentales de las personas (Low y Perner, 2012; Noyes y Christie, 2016), hacer juicios sobre la credibilidad del reporte de terceros (Kim, Kalish, Harris, 2012) y juzgar condiciones motivacionales que median interacciones humanas (Bascandziev y Harris, 2016; Marcovitch et al., 2015). En este, como en otros dominios específicos, el RI de los niños se fundamenta en su sensibilidad a las variaciones regulares de los eventos cotidianos (Opter y Bulloch, 2007). También usan el conocimiento de regularidades y patrones de comportamiento, para orientar sus acciones sobre el entorno (Waismeyer y Meltzoff, 2017). Incluso, repiten interacciones de agentes (presenciadas) para

generar cambios en las situaciones, e indagan por el acierto en sus conclusiones sobre covariación y causalidad (Gopnik et al., 2004; Sobel y Buchanan, 2009).

Una importante cantidad de estudios del RI, de la primera década del siglo XXI, según lo reportado por Kem y Jerm (2014), consideraron que la inferencia de patrones generales, a partir de la experiencia de unos pocos casos de la relación, es característica del desarrollo del conocimiento social. En este dominio en especial, las conclusiones generalizantes informan sobre las creencias de los niños respecto al origen de regularidades en su entorno (Gelman, et al., 2015). De hecho, errores en la generalización pueden ser consecuencia de sesgos en las creencias sobre representatividad de las categorías sociales (Noyes y Christie, 2016). Además, de sesgos en el conocimiento de propiedades distintivas de la apariencia de los miembros de la clase (Badger y Shapiro, 2012).

Resumiendo, si bien las generalizaciones inductivas, plausibles, son el fundamento de la generación de nuevo conocimiento, para este proceso los niños también disponen de otros recursos, como el conocimiento social implícito, por ejemplo (Evans, 2020). Diversos estudios con niños, en edades menores a 5 años, han probado el desarrollo temprano de la capacidad de hacer inferencias causales generalizantes, plausibles, a partir de conocimiento de patrones estadísticos y de probabilidades. Este proceso es posible en situaciones que presentan evidencia, suficiente, sobre la eficacia de los agentes causales, o en las que existe conocimiento previo de propiedades de las categorías implicadas (Gopnik et al, 2001; Sobel y Legare, 2014; Sutherland y Cimpian, 2017; Waldmann y Hagmayer, 2006).

Inferencia probabilística intuitiva

La evidencia empírica referida a la capacidad de los bebés para usar datos estadísticos en su RI, sugiere que la inferencia probabilística se desarrolla tempranamente. Este tipo de inferencia inductiva ocurre en función de orientar conductas, hacer predicciones, o elecciones y también ocurre en circunstancias de incertidumbre. Esta evidencia se sustenta, principalmente, en tres tipos de estudio: a) de la sensibilidad de los bebés a la frecuencia, regularidad y estructura estadística de los eventos, en diversos dominios; b) de la inferencia de relaciones causales, resultado de la observación de covariaciones y eventos probabilísticos; y c) de aprendizaje estadístico a partir de la inferencia de patrones probabilísticos, en dominios específicos.

Sensibilidad a patrones de propiedades y regularidades de los eventos

La evidencia antecedente ha demostrado que los bebés reconocen patrones de similitud entre atributos y relaciones de categorías, en una amplia variedad de representaciones de la información (Carey, 1985). También tienen memoria del funcionamiento y relaciones regulares entre instancias, agentes y artefactos (Bonawitz y Lombrozo, 2012). Así mismo, en los primeros años de vida se comprenden semejanzas en funciones y variables de sistemas de propiedades. Ese conocimiento puede ser usado para crear experiencias con los fenómenos y entidades del entorno (Lawson, 2017).

Dado que son sensibles a la similitud y regularidad estadística, los niños pueden inferir además de las características comunes de la superficie, o de la apariencia de los contenidos, las propiedades internas de los elementos. Incluso cuando los atributos percibidos se enfrentan con creencias sobre la pertenencia a una categoría conocida (Goswami, 2003, 2010). Estas capacidades se desarrollarán progresivamente, entre tanto el niño acceda a explicaciones de los hechos y aprenda a hacer analogías a partir de la experimentación de las cualidades físicas, e internas de las categorías. Es crítico, además, que tenga oportunidades para el descubrimiento de propiedades sensibles de los eventos, agentes y objetos (Wilkening y Cacchione, 2010).

Sensibilidad de los bebés a patrones estadísticos. Algunos logros tempranos en la inferencia de probabilidades se explican por la capacidad cerebral de los bebés para monitorear cambios producidos en las relaciones conocidas, durante ciertos periodos de tiempo. Así mismo, por la capacidad de detectar covariaciones y utilizar información sobre frecuencias y regularidades en la ocurrencia de eventos, para modificar su comportamiento (Cummins, 2004). Estas capacidades implican, por ejemplo, la sensibilidad de los bebés a frecuencias en la interacción social y a patrones de información estadística, explícita e implícita en la comunicación con ellos (Baillargeon et al, 2019).

Dichos patrones pueden presentar tanto similitud como diversidad y están disponibles para el razonamiento y aprendizaje cotidiano de los bebés. Sin embargo, el uso de tareas experimentales y contextos, altamente controlados, sigue siendo la alternativa para el estudio de la amplia variedad de respuestas explícitas del RI temprano. El control experimental es especialmente relevante para la investigación del desarrollo del pensamiento científico. Es usual que se indague el efecto del

conocimiento previo en la inferencia de probabilidades, mediante el uso de artefactos y tareas de laboratorio. Al respecto, se señala que este tipo de actividades distan mucho de representar fielmente las situaciones de RI en la cotidianidad de los niños (Leigthon, 2004).

Se considera que la sensibilidad de los bebés a patrones estadísticos visuales, facilita diversos procesos perceptivos y cognitivos. Como son: el reconocimiento de distribuciones espaciotemporales, la predicción de acciones, la inferencia de relaciones causales y el aprendizaje de categorías, entre otros (Tummeltshammer y Kirkham, 2013; Wu et al, 2011). Un ejemplo que ilustra estas capacidades es un estudio realizado por Xu y García (2008) sobre muestreo intuitivo. En cual se sugiere que los bebés, preverbales, pueden hacer predicciones de nuevos resultados, a partir de la observación de eventos aleatorios simples, y razonar de forma coherente con las probabilidades esperadas. En un experimento de seguimiento ocular, Tummeltshammer y Kirkham (2013) mostraron a bebés, de 8 meses, secuencias de eventos espaciotemporales que cambiaron de estado con tres probabilidades de transición. Los autores explican, que los estudios que incluyen el aprendizaje de eventos estadísticamente confiables y sus resultados, apoyan la tesis sobre la sensibilidad de los bebés a múltiples fuentes de variabilidad.

Posteriormente Deninson y Xu, (2014) plantearon que, en los experimentos de tiempo de mirada en muestreo de un evento único, los bebés podrían haber comprendido que mayor cantidad (abundancia, o más numerosos) es igual a más probable. Explican, que las respuestas de los niños pueden ser resultado del uso de heurísticos, producto de la capacidad de hacer comparaciones directas de cantidades de elementos presentes en una población, en vez de ser evidencia de inferencias fundamentadas en la comprensión de relaciones de proporcionalidad. Los resultados se describen como evidencia de que los bebés están en capacidad de hacer predicciones sobre la ubicación de un objeto deseable; en función de las proporciones de éste, en vez de guiarse por la comparación de cantidades observadas, o por la discriminación, entre muestras preferidas de las no preferidas.

Los hallazgos de Wu y cols., (2011) en un estudio sobre la detección de regularidades estadísticas entre eventos en el dominio visual, respaldan la idea de que la cognición de los bebés va más allá de la sensibilidad y detección de patrones estadísticos. Esto debido a que pueden aprovechar esa información para facilitar la categorización a partir de propiedades. De esta manera también lograrían nuevas predicciones, plausibles, sobre particularidades y comportamiento de los

objetos. Con apoyo de señales sociales, dichas predicciones favorecen el aprendizaje estadístico temprano. Por ende, se considera que la comprensión intuitiva de patrones probabilísticos en la interacción de variables del entorno físico, social y lingüístico (Gopnik, 2011; Schulz, 2012; Sobel y Legare 2014), subyace tanto a la anticipación (conocidas las frecuencias del evento), como a la predicción de nuevos eventos (sin experiencia, ni memoria de la ocurrencia).

Desde otra perspectiva, mediante el paradigma de máquina de lotería, Teglas y cols., (2011) investigaron si los bebés de 12 meses de edad tenían expectativas estadísticas (razonables) sobre un evento futuro. Para ello integraron conocimiento de cambios sistemáticos en la cantidad y disposición espaciotemporal de los objetos a la vistas, con información de la proporción de las muestras. Los investigadores afirmaron que los resultados son evidencia de la capacidad de los bebés para usar conocimiento relevante de dominio específico (de la Física en este caso) en la inferencia probabilística.

En la literatura especializada en diferentes enfoques del desarrollo cognitivo es frecuente encontrar precisiones como las descritas anteriormente. Lo cual sugirió que existe acuerdo entre los expertos respecto a las situaciones que favorecen el estudio del RI normativo. Por lo tanto, para indagar la inferencia de los niños se propone representar adecuadamente los patrones estadísticos, usando frecuencias naturales, por ejemplo (Zhu y Gigerenzer, 2006). Además, por medio de métodos experimentales, coherentes con el concepto formal de Probabilidad, atendiendo al uso que los razonadores dan a los heurísticos de representatividad. Puesto que esto facilita, o limita, la observación de mayorías en las muestras (Bryant y Nunes, 2012; Deninson y Xu, 2014).

Esta diferenciación es importante, dado que el RI sobre relaciones de proporcionalidad subyace a la inferencia probabilística y a la comprensión adecuada del concepto de Probabilidad (Howe, Nunes y Bryant, 2011). En cambio, el uso de heurísticos de representatividad, como guía en la predicción, se ha vinculado con sesgos, o fallas en el razonamiento probabilístico (Kahneman et al., 1982). Esto se debe, a que genera inferencias correctas con menor frecuencia (Pillow, 2002; Nisbett et al, 1982). Para estudiar el desarrollo temprano de la inferencia probabilística eficaz es importante tener claridad sobre las diferencias en el RI sobre proporciones. Esto implica considerar, en la creación de contenidos para el estudio de la inferencia de niños, la naturaleza del cálculo de probabilidades, así como la funcionalidad de heurísticos de cantidad y de contigüidad espaciotemporal.

Inferencia de covariaciones y relaciones causales probabilísticas. Una de las primeras investigaciones contemporáneas de la inferencia causal, de los niños en edad preescolar, fue de Gopnik, Sobel, Schulz y Glymour, en 2001. Este grupo estudió, desde una perspectiva de aprendizaje asociativo, la capacidad de inferir relaciones causales por observación de evidencia sobre patrones de variación y covariación. Los resultados confirmaron la hipótesis de los investigadores respecto la capacidad de integrar, eficientemente, conocimiento específico de dominio con información de probabilidad dependiente e independiente.

Recientemente, Waismeyer, Meltzoff y Gopnik (2015) indagaron el aprendizaje de relaciones causales, con la participación de niños de 24 meses. Los participantes observaron la ocurrencia de patrones probabilísticos de evidencia contingente con acciones de agentes, los cuales fallaron en algunos ensayos. Los resultados demostraron que los niños menores de 6 años aprenden acerca de las relaciones causales observando eventos aleatorios. También se verificó, que usan el conocimiento de causalidad para dirigir acciones hacia los objetos que mostraron más probabilidad de generar el efecto.

Posteriormente, Waismeyer y Meltzoff, (2017) replicaron el experimento, eliminando el contacto del agente con los objetos, puesto que los niños tienen un fuerte sesgo hacia la causalidad de contacto. En esa oportunidad, agregaron tareas para el aprendizaje de relaciones causales, entre eventos sociales, manteniendo la observación de resultados de manera probabilística. En conjunto, los hallazgos se interpretaron como evidencia de que los niños, menores de 6 años, aprenden relaciones de causa-efecto; razonando sobre relaciones deterministas. Así mismo, aprenden sobre patrones de evidencia de naturaleza probabilística. Estos resultados, apoyan la tesis sobre el desarrollo temprano de la habilidad para tomar decisiones y planificar la acción en condiciones de incertidumbre, en un conjunto amplio de situaciones y contextos.

En síntesis, diferentes perspectivas del razonamiento y del desarrollo del pensamiento científico coinciden en que, la predicción adecuada de la probabilidad de que múltiples causas afecten un resultado, mejora con el conocimiento de evidencia empírica sobre covarianza (Kuhn et al., 2008; Sternberg, 2010). Además, con el conocimiento teórico (que implica comprensión teórica) de las estructuras causales en dominios y contextos específicos (Kuhn y Dean, 2004). Lo anterior significa que, en condiciones de incertidumbre, los razonadores que disponen de estos

tipos de conocimiento harán predicciones más eficientes sobre relaciones de sistemas causales (Sobel et al, 2017).

Inferencia probabilística intuitiva y aprendizaje estadístico temprano. Desde la perspectiva neurobiológica de la cognición, el aprendizaje estadístico se define como el proceso mediante el cual las personas, desde la infancia temprana, detectan y extraen patrones confiables de información sobre los eventos cotidianos. Por ejemplo, razonando a partir de regularidades estadísticas en el espacio y el tiempo, en dominios como el lenguaje y el medio visual (Gómez, 2017; Hasson, 2017). En ese marco se ha investigado la capacidad de los bebés para detectar cambios en patrones estadísticos, mediante la manipulación de secuencias temporales o de matrices espaciales (Marcovitch y Lewkowicz, 2009). Recientemente, se estudió la flexibilidad en la percepción de la probabilidad de transición (cambio de estado durante un periodo de tiempo) de secuencias de información visual, hacia atrás o hacia adelante (Tummeltshammer et al., 2016).

Los bebés demuestran capacidad de codificar y representar unidades de información estadística relevante, entre otra irrelevante. Slone y Johnson (2018) investigaron aprendizaje estadístico visual usando el modelo de fragmentación. Este modelo plantea que la sensibilidad a las probabilidades de transición son subproductos de procesos como el aprendizaje asociativo, atención y memoria, en vez de ser resultado de la representación de propiedades de las relaciones estadísticas. A partir de los resultados, los investigadores sugieren que los bebés aprendieron fragmentos de información estadísticamente coherentes. Sugieren que, las representaciones que resultaron de este aprendizaje (estadístico visual) se explican mejor mediante modelos de representación de fragmentos de información, que mediante modelos de representación de relaciones estadísticas entre elementos.

Sintetizando, en el estudio contemporáneo del vínculo entre la inferencia probabilística intuitiva y el aprendizaje estadístico en la infancia temprana ha sido relevante la representación de relaciones y propiedades estadísticas de los eventos. Así mismo, representaciones de fenómenos como: frecuencia, homogeneidad, probabilidades condicionales y probabilidades de transición (temporal y espacial). Para el logro de los objetivos de esta disertación resultaron especialmente relevantes, hallazgos empíricos de los investigadores de la inferencia probabilística temprana. Puesto que, demuestran la relación entre la estructura de las tareas inductivas y el logro inferencial.

El RI eficaz en la infancia

Desde la perspectiva de la racionalidad ecológica (Gigerenzer, 2015) las conclusiones razonables son producto de la eficacia de los juicios y ocurren como consecuencia de la revisión de la racionalidad de las creencias. Desde el enfoque de los modelos bayesianos se estudia la racionalidad como una relación entre mente y entorno. De manera que, los niños forman creencias a partir de sus interacciones cotidianas y aprenden principios de razonamiento para verificar, por si mismos, la probabilidad de que éstas sean de utilidad para decidir el comportamiento (Culbertson et al., 2016).

De hecho, al inicio de la escolarización, los niños usan datos y patrones de información estadísticos con fines similares a los de la indagación científica. Estos propósitos son, básicamente, generar principios y explicaciones que permitan *predecir eventos en dominios específicos* (Kinnear y Clark, 2014). Las afirmaciones anteriores son respaldadas, además, por evidencia sobre la capacidad de los niños, menores de 6 años, para razonar eficazmente, sobre leyes y principios científicos. Incluso, en condiciones de incertidumbre por desconocimiento de conceptos y explicaciones científicas (Kinnear y Clark, 2014; Lane y Shafto, 2017; Samarapungavan, Mantzicopoulos y Patrick, 2008; Setoh et al, 2013; Shafto et al, 2008).

Desde la perspectiva metodológica, parece haber acuerdo entre los expertos en probar, o incluso defender la fortaleza de las inferencias logradas a partir de evidencia diversa, en comparación con las obtenidas a partir de evidencia menos diversa (Blok, Osherson y Medin, 2007). Al respecto, se ha propuesto que capacidad de inferir conclusiones plausibles en relación con la evidencia, es precursora de la habilidad para coordinar teoría y evidencia (Gopnik, 2012). Esta habilidad es considerada por Kuhn (2004, 2005, 2010b) una característica distintiva del pensamiento científico.

Atributos característicos de la inferencia inductiva eficaz

La precisión en la inferencia de propiedades y relaciones causales, en múltiples dominios es un atributo del RI explicado por la disponibilidad de teorías intuitivas coherentes. Los razonadores usan representaciones de los patrones de funcionamiento de propiedades causales generales, de sus teorías intuitivas, para inferir la relevancia de las variables involucradas en los contenidos de la situación (Bonawitz y Lombrozo, 2012). Incluso, logran este tipo de inferencia en

ausencia de evidencia directa, o contenidos explícitos que informen de la eficacia de las variables en los sistemas de causalidad (Johnson y Ahn, 2017). De la misma forma, estas teorías ponen a disposición del RI, representaciones de la organización y propósitos de la actividad mental que sustentan inferencias sobre agencia humana (Kominsky, Zamm y Keil, 2017).

La coherencia de la inferencia se explica por la capacidad de valorar la plausibilidad de las hipótesis logradas tras la consideración de evidencia (Pillow, 2002; Zimmerman, 2000, 2007). En algunos casos, la coherencia depende de generalizar resultados a partir de la predicción de varianza en las propiedades de las categorías, de forma que las conclusiones son progresivamente más razonables (Brandone, 2017; Sutherland y Cimpian, 2017). Dependiendo de la situación razonada, las inferencias coherentes se logran usando conocimiento implícito, que el niño recuerda espontáneamente. Este conocimiento apoya supuestos, más o menos razonables, respecto a la variabilidad de las relaciones, en condiciones diferentes a las de la situación presente (Diesnes y Perner, 1999). Esas condiciones generan incertidumbre y permiten valorar la coherencia de las representaciones conceptuales de los razonadores (Gelman y Davidson, 2013). Así, las inferencias coherentes consideran la posibilidad de verificar las relaciones que son determinantes en el resultado objeto de razonamiento, a futuro (Fischbein, 1975)

La relevancia es un principio que el razonador podría tomar en cuenta para integrar nueva información en la conclusión. Usando para este fin, el conocimiento de categorías o de causalidad que asocia con los contenidos de la situación razonada (Medin et al., 2003). Este principio también se refiere a la relevancia de la información que adicionan los niños a los contenidos de la situación. Dado que el razonador confía en la pertinencia de aportar nueva información, para llegar a conclusiones correctas, la relevancia se constituye en un atributo medible del RI. Porque, da cuenta de la relevancia de sus representaciones, o teorías abstractas, sobre categorías y relaciones de causalidad (Osherson et al., 1990).

De manera que, para que sea necesario aportar información relevante, las tareas deben incluir evidencias indirectas, nuevas propiedades y otro tipo de elementos que resulten desconocidos para los niños. La relevancia del conocimiento conceptual (y subjetivo) usado en la inferencia también atribuye eficacia al RI (Lawson, 2017; Hayes, 2007; Rhodes y Liebenson, 2015). Así mismo, las creencias relevantes respecto a los contenidos de dominio específico, modifican la eficacia de la conclusión (Kim et al., 2012; Weber y Osherson, 2010)

Por las razones descritas, la evaluación y medición de la eficacia en el RI tiene utilidad diagnóstica y es de interés para la Ciencia Cognitiva y la Educación. Así mismo, para los enfoques pedagógicos centrados en la formación de habilidades para vida en el siglo XXI. Estos consideran prioritario, entrenar a los aprendices en la inferencia de principios y estrategias de acción para abordar soluciones sustentables a los problemas que enfrenta la sociedad. Dichas soluciones deberán ser fundamentadas en la revisión del conocimiento del “reservorio histórico” de la Ciencia (Bouwmeester y Sijtsma, 2004; Moore et al., 2018). Por lo tanto, la formación de estas habilidades es objetivo de la Educación STEM, a nivel global (Zheng, Shu y Wong, 2018).

En síntesis, es posible afirmar que la medición de la eficacia en la inferencia inductiva, para la caracterización ecológicamente válida del RI, debe ser coherente con las conceptualizaciones de habilidades inferenciales tempranas. Se debe dar prioridad a los contenidos que evalúen capacidades para la búsqueda de conocimiento, la comprensión y la evaluación de evidencia e información verificable (Heit, 2000; Waldmann y Hagmayer, 2006). Evitando, por ejemplo, la evaluación de la eficiencia de los cálculos de probabilidad de niños de Preescolar. Dado que la evaluación justa de estas habilidades depende tanto de la edad del niño, de las oportunidades de aprendizaje estadístico, como de la coherencia de los criterios conceptuales para el diseño de contenidos (Saffran y Kirkham, 2018).

Medición de atributos de la eficacia en el RI

La investigación empírica del desarrollo del pensamiento científico ha contribuido a establecer las características del RI eficaz en contextos de incertidumbre. Se ha delimitado el tipo de conocimiento y estrategias inferenciales que mejoran la probabilidad y justifican las conclusiones. En este marco teleológico, se explica la importancia de incluir datos estadísticos en la explicación causal. Así como la necesidad de adecuar de contenidos y situaciones de dominio específico, con el fin de valorar la probabilidad de eventos (Ahl y Keil, 2017; Köksal-Tuncer et al, 2018; Kuhn, 2010b; Legare, 2014).

En contextos de causalidad probabilística, la posibilidad de evaluar o medir atributos de la eficacia en el RI depende de la capacidad del razonador para inferir reglas abstractas. Porque estas son el elemento de cohesión para la interpretación de nueva evidencia y de resultados estocásticos. De manera que, la inferencia inicial de reglas de relación genera un recurso cognitivo para resolver

situaciones problemáticas posteriores. El conocimiento de estas reglas es suma importancia para tomar decisiones acertadas y aprender conocimiento conceptual de dominios específicos (Nisbett et al, 1982).

A partir de las anteriores consideraciones se han creado indicadores para estimar la plausibilidad de las inferencias generalizantes y los atributos presión, coherencia y relevancia de las conclusiones. Dichas estimaciones suelen hacerse en diferentes modalidades y con actividades altamente estructuradas, en contextos a los que los niños, probablemente, no acceden en su cotidianidad (Schauble, 1996). Por estas razones, resultó pertinente para los fines de la presente disertación, adaptar y crear nuevos criterios de evaluación de la eficacia del RI. Para ser usados como fundamento de indicadores de desempeño, en situaciones con estructura inductiva multifacética. Estos criterios son:

a) generalización, plausible, de reglas de relación, covariación y propiedades distintivas de los sistemas de causalidad.

b) uso de conocimiento conceptual *preciso*; sobre categorías, mecanismos o principios que regulan relaciones de causalidad, patrones de dependencia e independencia y covariaciones sistemáticas en eventos, entre los que se incluye comportamientos intencionales de agentes.

c) delimitación *coherente* de las conclusiones, estableciendo la coincidencia entre estas y la nueva evidencia directa. Implica la comparación de nuevas representaciones de propiedades distintivas de las categorías, con conocimientos sobre regularidades en las relaciones y con evidencia conocida.

d) aporte de nueva información *relevante* para el control de la incertidumbre, se supone que las teorías intuitivas y las creencias de los niños son la fuente de información útil en la elección de alguna de las hipótesis que tienen probabilidad de ser plausibles, dada la evidencia directa e indirecta.

Evaluación de inferencias eficaces en múltiples modalidades. Las modalidades explícitas de la inferencia inductiva son conclusiones que se expresan verbalmente, como: juicios generalizantes, hipótesis, explicaciones y predicciones. Por esta razón, la conclusión ha sido la unidad primaria de análisis en la creación de contenidos para evaluación empírica de atributos del RI. Estas modalidades se han conceptualizado como una síntesis de los procesos cognitivos, implícitos, que les subyacen (Tomasello, 2010).

En este sentido, Gennari et al., (2002) sugirieron que el rendimiento en las tareas cognitivas experimentales, que incluyen entrenamiento para la producción de expresiones semánticas, es afectado por el lenguaje hablado. Explican que, la evaluación cognitiva por medio de tareas de razonamiento con la participación de niños, el estado de desarrollo de su lenguaje influye en procesos de origen cognitivo. Procesos como la focalización de la atención en las particularidades de la tarea y la expresión de juicios de similitud no lingüística. Por lo tanto, las tareas de RI dirigidas a niños en primera infancia resultaran más favorables para la inferencia, si incluyen estímulos visuales y de diversas características sensoriales (Lakin y Gambrell, 2012). Además, la demanda de respuestas explícitas deberá estar acorde con la capacidad del niño para comprender el lenguaje verbal y no verbal.

Probablemente, el aspecto semántico del conocimiento del niño en edad preescolar, es el que mejor representa un atributo mensurable de la inferencia inductiva, explícita. Esta característica, condiciona la tarea de razonamiento a la presentación de contenidos familiares para los niños. Aunque no es indispensable para la inferencia, que el razonador verbalice las denominaciones de los contenidos, o que nombre las categorías, si es importante que asocien información presentada en la situación con sus representaciones mentales, o con conceptos aprendidos. Es decir, los contenidos para la estimación de atributos del RI eficaz deberán ser, al menos parcialmente, análogos con elementos de algún un sistema de significados conocidos (Turiel, 2012; Holland et al, 1989; Nisbett y Ross, 1980).

Respecto a los aspectos pragmáticos de la medición de atributos del RI, también es importante considerar que, existe una relación de dependencia (que es directamente proporcional) entre la estimación de niveles de desempeño en las habilidades inferenciales y las demandas cognitivas de las tareas de RI (Vandekerckhove, 2014). Por ende, las tareas o situaciones para la indagación de la eficacia en las conclusiones, deben presentar evidencia de funcionamiento adecuado en un dominio específico y sus propiedades métricas deben demostrar porqué es adecuado (AERA, APA y NCME, 2014).

A continuación, se definen las formas explícitas de la conclusión que han sido indagadas para la evaluación y medición de atributos del RI:

Categorización. Es el proceso subyacente a la inferencia inductiva generalizante. Aunque, algunas líneas de investigación, no incluyen la categorización como modalidad del RI, se ha aclarado la importancia de la identificación de propiedades distintivas de las categorías. Dado que, a partir de este conocimiento el niño crea representaciones semánticas o conceptuales de las propiedades, elementos y sus relaciones. Diversas fuentes de evidencia sugieren que el conocimiento conceptual de categorías se hace explícito mediante la verbalización de sus denominaciones, o nombres (etiquetas) (Nguyen, 2012; Palmquist, Vikram y Jaswal, 2015). Se ha encontrado, por ejemplo, que los niños usan palabras precisas para describir categorías sociales desde los 2 años (Lane y Shafto, 2017)

Generalización. En este tipo de juicio se hace referencia explícita a la identificación de patrones de información relevantes para la inclusión de nuevos casos, representativos del evento o la categoría razonada. Se ha demostrado, ampliamente, que los niños en edad preescolar (entre 4 y 6 años), generalizan principios, reglas y elementos distintivos del sistema de relaciones razonadas (Bramley et al., 2018). Por lo tanto, la *conclusión* explícita es evidencia de la capacidad de discriminar casos en los que aplica la regla, o los generalizables, de los casos en los que no aplica, las excepciones (Lawson, 2017).

Hipótesis. La expresión de hipótesis plausibles es evidencia de la habilidad para representar probabilidades de relación entre variables o elementos del sistema; cuando no es obvio el vínculo causal entre estos, pero se conocen resultados sistemáticos de la asociación, o sus covariaciones y regularidades (Lane y Shafto, 2017). La expresión de ideas para apoyar a una hipótesis o refutar afirmaciones que la contradicen, también indica desarrollo temprano de un tipo de habilidad, de prueba de hipótesis, subyacente al razonamiento científico (Köksal-Tuncer y Sodian, 2018)

Explicación. La explicación aclara las relaciones al restringir y guiar la interpretación de información, y amplía las inferencias que pueden servir de base para nuevas predicciones e hipótesis (Legare, 2014). En contextos de RI sobre causalidad, la explicación plausible se caracteriza por basarse en la inferencia de causas ordinarias más que extraordinarias (Bonawitz y Lombrozo, 2012; Lane et al, 2016).

Predicción. Las predicciones plausibles indican que los supuestos del niño (sus modelos mentales) coinciden con las regularidades y probabilidades de eventos futuros, por ende, su

razonamiento es coherente (Nisbett y Ross, 1986). La anticipación, de eventos o condiciones conocidas, podría ser base para de este tipo de conclusión, especialmente si las circunstancias de las situaciones cambian y se desconocen las consecuencias de ese cambio (Kinnear y Clark, 2014; Setoh et al., 2013).

Formación del RI eficaz para el desarrollo de habilidades inferenciales

Recientemente, se actualizó la descripción de las habilidades para el desarrollo, aprendizaje y transmisión de conocimiento científico durante el siglo XXI. Se consideraba que éstas abarcaban a las competencias básicas del aprendizaje digital, el pensamiento crítico y la resolución de problemas reales (Greiff et al., 2014). Desde 2020, estudios del Foro Económico Mundial, proponen: resolución de problemas, pensamiento creativo, perseverancia, autoeficacia, motivación, conciencia y metacognición. Probablemente, las consecuencias socioeconómicas de la Pandemia de Covid-19, motiven próximas redefiniciones. Inclusive, de los procesos educativos y culturales para la formación de las competencias cognitivas necesarias para crear unidad, igualdad y productividad en el mundo del siglo XXI (World Economic Forum, 2021)

Tradicionalmente, desde el modelo de la Educación STEM, se han descrito ventajas de investigar el desarrollo de las habilidades cognitivas implicadas en el aprendizaje y reproducción del conocimiento en diversos contextos (Mohlhenrich, Sergey y Spencer, 2018; Ricco, 2015). Dado que, dichas habilidades fomentan el rendimiento académico y se consideran el fundamento de la explicación científica de las sinergias y relaciones en un mundo tecnológicamente dinamizado (Kelly, Lesh y Baek, 2010; Granovskiy, 2018; MEN, 2012). De forma similar, el acuerdo internacional sobre la necesidad de transformar los currículos escolares al Modelo K-12, ha enfatizado el enfoque de la Educación STEM, atribuyendo gran importancia al aprendizaje de habilidades para razonar, eficientemente.

Respecto a la educación del RI eficaz, algunas perspectivas pedagógicas proponen estrategias innovadoras para formar el pensamiento. Aprovechando la capacidad de aplicar el conocimiento y aprender tareas de resolución de problemas abstractos y matemáticos. Por medio del arte, por ejemplo, y de otros recursos culturales disponibles. Con el fin de preparar a los jóvenes en la transferencia de habilidades para resolver problemas complejos y aportar conocimientos diversos, o interdisciplinarios (Kuhn, 2005; Singh, 2021). Organizaciones mundiales proponen crear pedagogías que consideren la relevancia de pensar y aprender a partir de diferentes tipos de

datos estadísticos. Diversas culturas valoran el conocimiento científico y de evidencias sobre factores de cambio en los sistemas y procesos causales inherentes a dominios específicos (Griffin, McGaw y Care, 2012; OECD, 2009, 2012a, 2013a; UNESCO, 2005, 2015, 2017).

Es frecuente, en la literatura sobre desarrollo cognitivo, que se proponga la mediación pedagógica para el aprendizaje eficiente de habilidades inferenciales. Se sugieren y se han creado métodos para generar evidencia empírica sobre la práctica cotidiana de estas habilidades, en el aula y fuera de escuela (Kuhn, 2010a; Hauze y French, 2018; National Science Foundation, 2007). Para lograr este propósito, algunas líneas de investigación proponen usar contenidos con datos sobre regularidades y situaciones probabilísticas. Dado que estas facilitan el diagnóstico de potenciales y dificultades en el razonamiento estadístico, incluso temprano en la infancia (Gopnik, 2012). Por ejemplo, una dificultad común en el proceso de inferencia estadística es subestimar el riesgo (Borovcnik y Kapadia, 2018; Gigerenzer, 2015). De manera que, existe la expectativa de mejoramiento en la habilidad de los adolescentes para evaluar críticamente el azar, la certeza y la incertidumbre, por efecto de la educación (Washer, 2007; Batanero et al 2016; Batanero y Chernoff, 2018)

Inferencia eficaz temprana y generación de conocimiento de áreas STEM

Inferencias inductivas eficaces han sido descritas como precursores del pensamiento científico, incluso las de niños menores de 6 años. Esto se explica porque el RI es también una extensión de habilidades inherentes a la indagación de cotidiana de eventos novedosos (Brenneman y Louro, 2008). Por estas razones, la práctica frecuente de inferencias mediante contenidos y herramientas de indagación científica, fomenta el cambio conceptual y la comprensión de principios de los dominios STEM (Gelman y Brenneman, 2012; Kelly et al., 2010).

La implementación de estas innovaciones pedagógicas, en la escuela, podría facilitarse mediante la apropiación de instrumentos diagnósticos de las habilidades inferenciales. Así mismo, por medio de la adecuación pedagógica de tecnologías que permitan a los educadores *obtener datos a partir de las inferencias de los niños*. Además, se sugieren usos pedagógicos concretos para ese tipo de datos, con los siguientes objetivos:

- Identificar la *agencia* y la igualdad de las personas para participar en la construcción científica (Epstein y Fischer, 2017; Ricco, 2015)

- Comprender *relaciones causales y covariaciones*, infiriendo conocimientos básicos que los preparan para la enseñanza posterior mediante experimentos (Siler, Klahr y Price, 2012)
- Motivar la búsqueda del *conocimiento científico* en una variedad amplia de contextos (Gelman y Brenneman, 2012; Brenneman y Louro, 2008).
- Innovar las prácticas de *formación de maestros*, incluyendo la colaboración e integrando tecnología directamente en el proceso de autoformación (Herro y Quigley, 2016; Moore, 2009).

Actualmente, los estudios de los factores relevantes en el desarrollo y transmisión de habilidades cognitivas para la productividad en la sociedad del siglo XXI, sugieren el uso de herramientas de evaluación diagnóstica (Culbertson, 2016). Se propone la realización de estudios empíricos que midan esos factores y sus procesos determinantes. Por ejemplo, refieren la importancia de indagar la influencia del apoyo familiar para adquirir habilidades tecnológicas (van Laar et al., 2020). Se considera, ampliamente, que dichos procesos se asocian con prácticas de formación del pensamiento, tanto académicas como sociales (Figura 1).

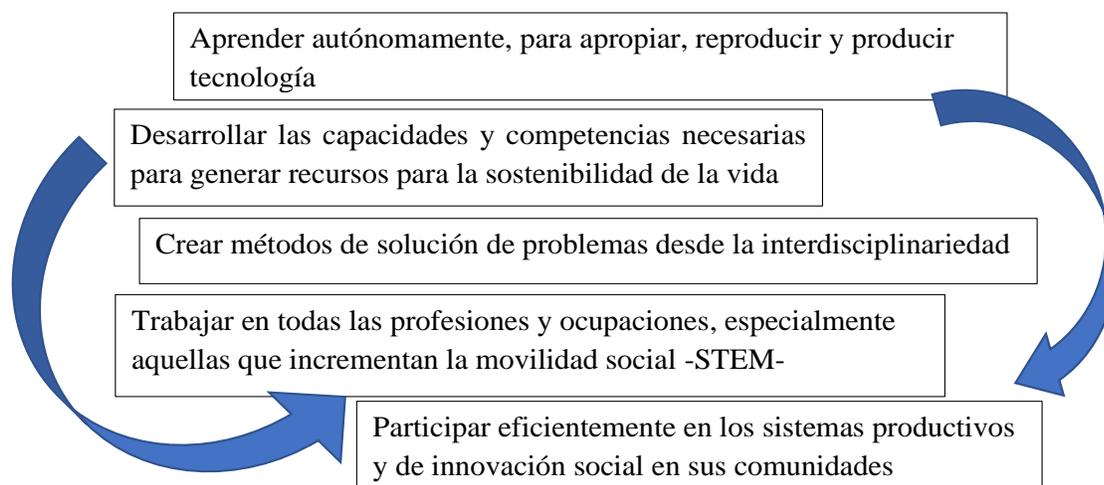


Figura 1. Ciclo de la formación de pensamiento para la transformación del conocimiento STEM, en todos los niveles de la Educación y contextos sociales. Fuente: elaboración propia, a partir del marco teleológico de la Educación STEM.

Desafíos para la formación de habilidades de inferencia en Preescolar. Teóricamente, las herramientas diagnósticas y de evaluación del desarrollo cognitivo son recursos disponibles para los agentes de la acción pedagógica en Preescolar. Se dispone de evidencia y antecedentes teóricos suficientes para sustentar que los educadores de Preescolar están involucrados en el desarrollo de las habilidades de pensamiento científico, tanto de los niños a quienes forma, como el propio (Kuhn, 2010a; Shtulman y Schulz 2008; Nancekivell y Friedman, 2017). A pesar de esto, faltan publicaciones de estudios empíricos del RI, en primera infancia, que hubieran obtenido datos de la observación y evaluación realizada por educadores, o familiares de los niños participantes.

Otra dificultad por abordar, en el campo de la formación del pensamiento infantil, es la escasez de reportes de medición de atributos de la inferencia inductiva de niños menores de 6 años, desde el marco de la Educación STEM. Incluso, entre los estudios de países que han adoptado el modelo, y no se encontraron publicaciones de informes de investigación empírica del RI de estudiantes del nivel Preescolar, en los países Latinoamericanos que han empezado la implementación del Programa para la Evaluación Internacional de Alumnos, PISA (OECD, 2015).

Esto puede ocurrir por las siguientes razones:

A) los estudios del RI se concentran en poblaciones y recursos del ámbito escolar y por lo general, incluyen estudiantes que demuestran competencia en la verbalización de sus conclusiones. Excepcionalmente se han obtenido datos de las prácticas de razonamiento sobre dominios STEM en las aulas, o fuera de los laboratorios de cognición (Avila y Barragan, 2018).

B) existen desacuerdos frente al desarrollo temprano de capacidades cognitivas involucradas en la inferencia inductiva eficaz (como las Funciones Ejecutivas, por ejemplo), entre modelos de aprendizaje y modelos de RI (Gelman et al, 2013; Gopnik, 2011, 2012).

C) las iniciativas que proponen crear contextos y estrategias para la indagación de habilidades involucradas en la enseñanza-aprendizaje de conocimientos estocásticos, ecológicamente válidos, se han realizado con estudiantes desde los últimos grados de Primaria (Eichler y Vogel, 2012; Frost et al, 2019; Gandhi, 2018).

Con el objetivo de asumir dichos retos se propone emprender innovaciones en las experiencias tempranas de formación de la cognición y actualizar objetivos de la Educación Inicial. Fomentando el interés de los agentes formadores, principalmente, en los siguientes ámbitos de la acción pedagógica:

- Presentación adecuada de datos estadísticos, de probabilidades y de relaciones inciertas en sistemas correlacionales y causales, para facilitar la inferencia de hipótesis y explicaciones sobre fenómenos en los dominios STEM.
- Innovación y creación de contenidos que sean semánticamente precisos, pertinentes en relación con el desarrollo del lenguaje, ecológicamente justificados y que favorezcan el RI eficaz de relaciones (determinadas y aleatorias) relevantes para abordar soluciones a problemas cotidianos.

¿Qué tipo de factores influyen los procesos cognitivos precursores de las habilidades del pensamiento científico?

Las preferencias culturales parecen ser determinantes en el desarrollo de procesos asociados al aprendizaje y transferencia de habilidades de pensamiento científico. Incluso se ha estudiado la influencia cultural en las estrategias de inferencia, estilo de pensamiento y el razonamiento cotidiano (Nisbett et al., 2001). No obstante, sigue vigente el debate respecto al tipo de prácticas que son críticas para el desarrollo del pensamiento científico (Dean y Kuhn, 2007; Strand-Cary y Klahr, 2008). Así mismo, la influencia cultural prepara a las personas, desde la infancia, para comprender que la disponibilidad de información y conocimiento de estadísticas, y la experticia, los facultan para transformar e intervenir resultados (Aguiar, Stoess y Taylor, 2012).

Dado que el contexto social del niño provee gran parte de la información relevante para evaluar probabilidades de acierto en la elección y los juicios generalizantes, éste se considera como una fuente importante de covariables para el estudio de los procesos de razonamiento sobre situaciones problemáticas de la cotidianidad (Morewedge y Kahneman, 2010). Es probable que, la preferencia de los adultos oriente la atención y el razonamiento de los niños, hacia cierto tipo de conocimientos, o particularidades de ciertas categorías. Influyendo así el desarrollo de singularidades en sus estrategias de inferencia (Sternberg, 2010).

La evidencia actual muestra distinciones categóricas en las prácticas familiares de formación del pensamiento entre grupos culturales diferentes. Nisbett et al. (1983) explican el acuerdo social respecto a que el RI es más preciso si se comprende la información estadística asociada a los eventos. Desde esta perspectiva, las prácticas familiares pueden maximizar tiempos y experiencias para pensar e indagar eventos reales, incluso para instruir sobre la evidencia y las

categorías que son relevantes para resolver problemas cotidianos (Schauble, 1996; Strand-Cary y Klahr, 2008).

La frecuencia de las *experiencias de indagación* de relaciones, variables y sistemas de causalidad probabilística es un factor importante en el desarrollo de conocimiento conceptual durante la infancia. De hecho, se han documentado cambios favorables en las estrategias de inferencia en función del tiempo de práctica (Kuhn, 2005; McGuinness Institute; 2015). La práctica prolongada, a mediano y largo plazo, mejora las habilidades cognitivas implicadas en el pensamiento científico (Kuhn, 2005; Kuhn, Schauble y Garcia-Mila, 1992).

Influencia social en la generación de representaciones conceptuales significativas.

Las interacciones sociales proveen información relevante para la formación de creencias y conceptos relacionados con la Ciencia. También promueven la representación de significados y modelos de explicación sobre la importancia del conocimiento valorado colectivamente (Kushir y Klein, 2018). Así mismo, la información crítica para inferir y representar la relevancia de las estructuras y explicar sistemas de relaciones sobre los que se razona cotidianamente (Ifenthaler y Seel, 2011; Osborne, 2013). De hecho, los niños aprenden modelos explicativos del valor de los principios científicos para la productividad en la sociedad, a partir de las interacciones familiares (Burris, 1982).

La información, evidencia y variables de origen social también influyen en la transición entre las teorías ingenuas y la representación de conceptos de dominios específicos, entre estos los de áreas STEM. Carey, (2009) planteó que dichas transiciones se dan por transferencia de conocimiento social. De manera que es probable, que el desarrollo y cambio conceptual sean afectados por el interés de los formadores en que los niños razonen sobre el conocimiento promovido socialmente, o el que se considera deseable para el aprendizaje (Gelman y Brenneman, 2012). En este proceso de intercambio social y transmisión de representaciones conceptuales se consolidan creencias sobre el conocimiento relevante para el avance social.

Las creencias sobre el valor de la evidencia como objeto de razonamiento sobre la Ciencia y otros cuerpos de conocimiento teóricos, pueden formarse, o sesgarse por influencia de las *suposiciones colectivas* (Kuhn y Dean, 2004). Por lo tanto, el interés en razonar sobre los conceptos y en usar el conocimiento de dominios STEM, ha sido justificado y promovido en diversos ámbitos

sociales y educativos, durante lo corrido del siglo XXI. Esa tendencia se asocia también, con la comprensión de que el conocimiento científico tiene relación con la implementación de soluciones a problemas globales, por lo que éste es aceptado interculturalmente (OECD, 2009, 2014; National Science Foundation, 2007).

Practica del RI a partir de hechos verificables en la cotidianidad. Para la investigación del desarrollo de habilidades cognitivas es importante preguntar por las prácticas de indagación de eventos cotidianos y las estrategias de inferencia que les dan sentido. Puesto que, las explicaciones a dichos eventos suelen enmarcarse en un cuerpo de conocimiento de dominio específico (Ifenthaler y Seel, 2011). De hecho, dar sentido a la experiencia a partir de un marco conceptual, o usando conocimientos objetivos como referentes, es practicar ciencia implícitamente, según lo explica Deanna Kuhn (2003, 2010b).

Algunos acuerdos sobre la relevancia del RI en la formación de conceptos ha favorecido la tendencia a considerar que el aprendizaje del cambio conceptual, en las disciplinas STEM, depende de las oportunidades para abordar el tema de forma activa y crítica en las aulas. En consecuencia, se convirtió en prioridad pedagógica, promover una variedad de habilidades de razonamiento a partir de prácticas en las que se estudien principios de las teorías y evidencias (Kuhn, 2010; Marin y Halpern, 2011; Ricco, 2015). En ese sentido, Ricco (2015) plantea que el pensamiento e indagación de eventos cotidianos, en la infancia, son precursores de las habilidades inferenciales asociadas a la indagación científica.

Si bien el aula provee conocimiento, información, evidencias y variables para la práctica y experimentación con los contenidos de la Ciencia; el entorno primordial para las prácticas de indagación es el hogar, especialmente los espacios de interacción familiar. Es importante señalar, que la probabilidad de que la actividad social de los niños colombianos se restringiera a los entornos familiares, aumentó drásticamente durante los años 2020 y 2021, por efecto de la crisis sanitaria desencadenada por la Pandemia. Esta situación ofreció la oportunidad de ampliar el estudio de la diversidad de contextos considerados marcos relacionales y sociales primordiales para fomentar el desarrollo de habilidades cognitivas (Damon, 1990).

Prácticas de socialización de la cognición. Mediante las prácticas cotidianas de socialización de la cognición, las generaciones precedentes reproducen un acervo común de

conocimiento, que es fundamental para el desarrollo cognitivo la infancia temprana. La apropiación, transmisión y reproducción de conocimientos socialmente aceptados, son factores esenciales del proceso de socialización de la cognición (Burris, 1982; Nisbett et al., 2001;). De manera que este proceso está implicado en la formación del pensamiento en la cotidianidad, así como en la interacción pedagógica y la instrucción formal (Bridges et al., 2015; Ferrar et al., 2019).

Las influencias culturales pueden explicar diferencias en los intereses de socialización de la cognición y en las prácticas que los concretan, como explicaciones, búsqueda y omisión de evidencia e información (entre otras). Estas influencias se han estudiado en situaciones de razonamiento sobre causalidad, en las que las explicaciones que los adultos brindan a los niños demuestran preferencias por conocimiento de dominios específicos, por ejemplo, en el dominio de la Biología y de la Psicología (Erickson & Lockhart, 2010).

Respecto a las diferencias individuales en las habilidades cognitivas, asociadas a la diversidad en las prácticas de socialización de la cognición, los datos son poco concluyentes. Es poco frecuente que los estudios del desarrollo cognitivo, en la infancia temprana, destaquen la formación del pensamiento, especialmente si participan familias Latinas. Por ejemplo, Bridges et al. (2015) indagaron el tiempo de práctica de actividades domésticas y el compromiso cognitivo de los niños en ellas. La influencia del tipo de práctica de socialización en la cognición de los niños se asoció con rendimiento escolar de la madre, de forma más significativa que con la reproducción de conocimiento social.

Algo similar sucede con los estudios empíricos del desarrollo cognitivo en Colombia. Se encontró que, en ninguno se indagó la influencia de las prácticas de socialización en las diferencias en el RI de los niños participantes. Sin embargo, se ha relacionado el desempeño adecuado de tareas de RI con el desarrollo del lenguaje (Duque et al., 2010; Chavez y Flores, 2011). Es apropiado señalar que este tipo de asociación limita la comprensión de la naturaleza no lingüística del conocimiento (Waxman, 2010). Por otra parte, en un estudio reciente Cerchiaro y Puche (2018) consideraron relevante la influencia de las claves lingüísticas y sociales en las inferencias inductivas de niños colombianos, de 2 años de edad.

Supuestos teóricos que justifican la propuesta metodológica para la caracterización del RI temprano

Para finalizar la revisión de antecedentes se sintetizan los supuestos teóricos que sustentan el problema y justifican las preguntas e hipótesis de investigación. En la primera parte de la revisión se describen posibles orígenes de los principios normativos del RI plausible en la infancia temprana, así: a) sensibilidad a patrones de información estadística y aprendizaje estadístico temprano; b) sensibilidad a la similitud y diversidad de las propiedades de categorías familiares; y c) sensibilidad al tamaño y composición de muestras diversas. A partir de esta evidencia, se supone que los niños hacen inferencias plausibles en contextos de causalidad probabilística antes de los 4 años.

Sin embargo, es escasa la evidencia de diferencias individuales en diversas modalidades explícitas de la inferencia generalizante y del funcionamiento de las tareas en contextos cotidianos, como el aula de Preescolar. Es particularmente notoria la falta de estudios del RI de niños con edades entre 4 y 6, pertenecientes a poblaciones con altos índices de pobreza y oportunidades limitadas de acceso a diversas fuentes de conocimiento.

Probablemente, se requieren nuevas metodologías, contenidos de dominio específico, e indicadores de desempeño, que faciliten la delimitación de las características distintivas del RI normativo temprano. Además, son necesarios instrumentos de indagación que integren saberes, de los expertos en evaluación cognitiva, en los dominios del contenido y en la formación de la cognición de los niños de Preescolar. Para enfocar estas suposiciones se plantearon las siguientes preguntas:

1. ¿podrían estudiarse las diferentes modalidades explícitas de la inferencia generalizante mediante situaciones multifacéticas?, para lograr este objetivo
2. ¿resultaría pertinente usar contenidos de dominios STEM?, y de ser pertinente,
3. ¿podrían obtenerse datos y evidencia de esta capacidad en el aula de primer grado de Preescolar?

Posteriormente se sustenta que, durante el RI los niños usan y transforman información estadística, además de conocimiento semántico, abstracto y conceptual, generando nuevo conocimiento, verificable. Los estudios empíricos de desempeño en la inferencia inductiva, usualmente, crean criterios de comparación de respuestas del razonador con principios conceptuales y pragmáticos de diferentes dominios. Estos criterios se usan para delimitar un conjunto determinado de conclusiones que son más justificables que otras. Son abundantes las

situaciones de RI con contenidos de: sociales, Psicología y Biología. Incluso, se han usado situaciones que combinan contenidos de estos dominios e información ambigua, para crear incertidumbre y motivar el uso de representaciones de reglas de relación abstractas.

Desde la perspectiva metodológica, está claro que es posible caracterizar atributos de las inferencias eficaces. No obstante, en gran parte de los estudios publicados, éstos se han evaluado por separado y con la participación de razonadores mayores de 5 años, escolarizados. Probablemente, la mayor dificultad para la caracterización del RI en la infancia, es el desacuerdo respecto a la edad en la que los niños logran inferencias eficaces con mayor frecuencia, y si lo hacen en diferentes dominios STEM. Se consideró razonable que, el avance en esta discusión depende de generar evidencia empírica de los factores que favorecen la coincidencia entre la conclusión y los indicadores de habilidad en el atributo estudiado. Indagando si las diferencias de edad son uno de esos factores críticos.

Respecto a esta alternativa, se propone la creación de indicadores de desempeño, de acuerdo con principios del dominio del contenido razonado y datos provenientes de la observación de respuestas de niños en el rango de edad de interés. Una vez obtenidos los indicadores es viable integrarlos a una herramienta de evaluación y probarlos en diferentes contextos de la actividad cognitiva. En ese orden de ideas, se diseñó un instrumento de medición de la eficacia en el RI de los niños, llamado IMERIN. La propuesta se orientó a partir de las siguientes preguntas:

4 ¿Qué tipo de conclusiones indican precisión, coherencia y relevancia en la inferencia inductiva de los niños, en dominios STEM?

5 ¿Se podrían medir múltiples atributos de la eficacia en las inferencias de niños, menores de 6 años, en una misma actividad de RI?

La literatura sobre RI del siglo XXI, destaca el acuerdo teórico en que, la inferencia de probabilidades es un mecanismo subyacente al RI eficaz y es influida culturalmente. Las creencias son uno de los factores de influencia en el RI mejor estudiados, así como la focalización de la atención en ciertas clases de evidencia, propiedades de las categorías y uso de claves lingüísticas, o de términos genéricos, por ejemplo. Estas ideas han movilizad la creación de métodos ecológicamente confiables y válidos para el estudio del desarrollo de habilidades cognitivas inherentes al pensamiento científico, en la adolescencia.

Con el fin de aportar a este objetivo se enfocó la medición, en la capacidad de los niños para hacer generalizaciones plausibles y en la pertinencia, suficiencia y objetividad de los

indicadores de eficacia en la inferencia. Respecto a la relación entre las creencias y el RI se planteó la siguiente hipótesis: las creencias sesgadas pueden influir la irrelevancia de las conclusiones respecto a la evidencia, de igual forma, las creencias coherentes con los principios del dominio de los contenidos, influyen la probabilidad de las conclusiones. Esta influencia puede probarse usando evidencia directa e indirecta y en condiciones de la situación parcialmente inciertas. Para probar esta hipótesis, se debería responder a una pregunta común en el debate frente al desarrollo temprano de la capacidad de coordinar teoría y evidencia, para el caso propuesto la cuestión se presenta así:

6 ¿los niños con edades entre 4 y 6 años, están en capacidad de resolver actividades de RI con múltiples situaciones, coordinando sus conocimientos y las evidencias, de forma eficaz?

Algunos planteamientos de la Psicología Social y de la teoría sociocultural vinculan la intencionalidad, e intereses intelectuales de los adultos, con el desarrollo cognitivo de los niños a quienes forman. En ese sentido, se plantearon hipótesis respecto a posibles relaciones entre el conocimiento deseable para el aprendizaje y diferencias individuales en el desarrollo de habilidades inferenciales en la infancia temprana. En la indagación de estas relaciones se consideró que, probablemente, las prácticas de socialización de la cognición demuestran preferencias por comunicar, buscar o enseñar conocimiento de dominios STEM.

El proceso de indagación de los supuestos descritos se orientó por las siguientes preguntas:

7 ¿las diferencias interindividuales en los desempeños de las actividades de RI se asocian con la frecuencia de las prácticas de formación de cognición en la familia?

8. las madres o adultos involucrados en la formación de los niños, ¿promueven prácticas de razonamiento sobre el conocimiento de dominios STEM?

La revisión de estudios del desarrollo, aprendizaje y trasmisión de habilidades inferenciales precursoras del pensamiento científico, indica que la Psicología y la Educación inicial podrían requerir modelos de diagnóstico cognitivo, actualizados. Así como métodos para la caracterización del RI de los niños, ecológicamente válidos y confiables. Se asume que dichos métodos aportarían evidencia relevante para la toma de decisiones sobre cambios curriculares e innovación pedagógica, en Preescolar.

9 ¿es probable que el conocimiento de las características distintivas del RI de los niños colombianos influya la toma de decisiones, para la formación de habilidades de pensamiento científico, desde Preescolar?

A continuación, en la Tabla 1, se presenta la síntesis de las preguntas e hipótesis que orientaron la investigación. Estas son inherentes a los casos teóricos que fundamentaron la metodología y están justificados por los resultados de la revisión de antecedentes.

Tabla 1.

Resumen de preguntas e hipótesis que orientaron los objetivos de cada estudio y etapas del trabajo metodológico

	Justificación	Objetivo de estudios	Referentes normativos	Componentes indagados
Estudio 1	¿Instrumentalizar permite resolver el problema de caracterizar el RI mediante medición de atributos de la eficacia en la inferencia?	Construir los contenidos de los instrumentos con los expertos temáticos Evaluación de criterios conceptuales para crear indicadores de eficacia RI en dominios STEM	Estructura inductiva es definida como la macroestructura del proceso cognitivo del RI Atributos de la inferencia componen la microestructura del RI	Mecanismos de inferencia eficaz en dominios específicos Modalidades explícitas de la conclusión en edad preescolar
Estudio 1-b	¿El estudio es viable dadas las restricciones al contacto físico y social durante la Pandemia por Covid -19?	Establecer la viabilidad de la aplicación, a distancia, de instrumentos para caracterizar el RI, con uso de TIC	Mediación del conocimiento social en el desarrollo conceptual El conocimiento de Categorías- causalidad determina la eficacia del RI	Verbalización de la conclusión – acompañamiento del adulto Significados de formación de las habilidades para la vida del siglo XXI
Estudio 2	¿Los indicadores del instrumento permiten medir la eficacia del RI? El IMERIN, su resultado y la interpretación propuesta son válidos y confiables?	Demostrar la utilidad científica de los indicadores de eficacia e ítems del IMERIN y de los usos diagnósticos propuestos para el resultado	Validez de contenidos y validez externa Requerimientos internacionales sobre confiabilidad de la medida y de la interpretación del resultado	Cómo la instrumentalización de la evaluación permite caracterizar el RI Uso de herramientas estadísticas para probar hipótesis y verificar supuestos sobre la estimación de atributos RI
Estudio 3	La eficacia en el RI puede ser influida por la frecuencia de las prácticas de formación de la cognición, porque estas moderan el desarrollo conceptual	Indagar las prácticas de búsqueda de conocimiento socialmente aceptado Indagar posibles claves sociales, y pedagógicas que influyen la habilidad inferencial temprana	Validez ecológica-externa en contexto particular estudiado (en Pandemia) Diversidad y aprendizaje cultural influye el razonamiento eficaz durante toda la vida	Factores familiares moderadores de la búsqueda de conocimiento deseable para el aprendizaje en la infancia Variables individuales y su relación con hábitos de formación de la cognición

Capítulo 2: Método

Resumen de las etapas del proceso metodológico

La investigación se fundamentó en 3 tipos de estudios, realizados en 3, o 4 etapas cada uno. Estos se justificaron por la necesidad de sustentar empíricamente los argumentos del modelo cognitivo propuesto para la caracterización del RI (Leighton, 2007). Con este objetivo fueron considerados los estándares normativos para valorar la utilidad científica de los instrumentos de evaluación diagnóstica en Psicología y Educación, propuestos por las organizaciones American Educational Research Association (AERA), American Psychological Association (APA) y National Council on Measurement in Education (NCME), 2014.

Como innovación metodológica se destaca la inclusión, en cada una de las etapas, de los criterios de evaluación del razonamiento, o del desarrollo de la cognición, que usan las educadoras en primera infancia, los científicos y las madres. Puesto que todos ellos son formadores del pensamiento científico, en diferentes contextos y niveles de Educación. De esta forma se creó un precedente para el diseño e implementación de metodologías de evaluación de diversas habilidades inferenciales de los estudiantes de Preescolar, en el aula y fuera de ella. La integración e interacción multidisciplinaria también orientó la interpretación de conceptos y apreciaciones respecto a los propósitos de la Educación STEM para el siglo XXI.

Participantes

Durante el desarrollo de la investigación se contó con la participación de 264 personas, 148 niños y 116 adultos. En cada una de las descripciones de las etapas se especificarán las características de cada grupo de participantes, por las cuales se justificó convocar su participación. El muestreo para la participación fue por conveniencia.

Estudio 1- primer etapa. En la etapa 1A se crearon, adaptaron y evaluaron situaciones y contenidos de los instrumentos de medición de la eficacia en el RI y de indagación de frecuencias de las actividades de formación de la cognición en la familia. Se hicieron pruebas del funcionamiento de los criterios e indicadores de plausibilidad, para conformar y probar el funcionamiento de un instrumento de medición de atributos de la inferencia inductiva, en tres modalidades explícitas (intervención en un resultado, elección de evidencia y adición de información). Este estudio se fundamentó en los resultados de la Revisión Sistemática de la Literatura. Concluyó con la conformación de una muestra de 120 actividades de inferencia

inductiva, de las cuales se seleccionaron algunas para ser probadas en cada etapa del diseño de contenidos.

La etapa 1B tuvo por objetivo establecer la validez de los criterios para el diseño de indicadores de eficacia en la inferencia. Con este fin, los ítems fueron evaluados por profesores y científicos expertos en enseñanza de las ciencias, en los diferentes niveles de Educación. Los expertos evaluaron la pertinencia, suficiencia y relevancia de los conceptos y relaciones representadas en los contenidos, en relación con los principios científicos que las definen desde las áreas STEM. Por último, en la etapa 1C se determinó y probó el funcionamiento de la estructura inductiva de las 12 situaciones de RI elegidas. Se analizaron los resultados y se construyeron actividades de generalización, con los respectivos criterios de evaluación e indicadores de plausibilidad de la conclusión. Así, se conformó la versión inicial del instrumento de evaluación multidimensional del RI de los niños, IMERIN-v1.

Estudio 1 - segunda etapa. En la etapa 2A se adaptaron los contenidos del IMERIN-v1, para medir eficacia en la inferencia a partir de la verbalización de conclusiones y los indicadores fueron evaluados por profesoras de Preescolar. En la etapa 2B se realizó la aplicación de prueba para evaluar el funcionamiento de los nuevos ítems, además se indagó si las condiciones de aplicación, limitadas por restricciones al contacto físico y por el cierre de las aulas, tenían efectos relevantes en las respuestas explícitas de los participantes. Finalmente se conformó el IMERIN-P con 9 actividades, esta fue la versión de prueba de las condiciones de la aplicación a distancia.

En la etapa 2C se evaluó la precisión semántica de los indicadores de eficacia del IMERIN-P y de los ítems de la encuesta de socialización de la cognición, mediante la técnica llamada Redes Semánticas. Los resultados se usaron para analizar si los términos, palabras y frases tenían el significado atribuido en los instrumentos, en comparación con las palabras y significados que los expertos en STEM y las profesoras de Preescolar, creían que los niños asociarían a los contenidos. Es decir, se comprobó la pertinencia y coincidencia semántica de los indicadores de eficacia y los ítems del instrumento.

Estudio 2. Inicialmente, en la etapa 2A, se realizó un juicio de Psicólogos expertos, para establecer la validez de contenidos de las 9 actividades del IMERIN-P, compuestos por 36 ítems. Posteriormente, en la etapa 2B, se realizó la aplicación para validar el uso propuesto para los resultados del instrumento. Este proceso consistió en generar evidencia empírica de la confiabilidad

de la interpretación del puntaje sugerida y probar la utilidad de los niveles de desempeño en la indagación científica de las diferencias individuales en la eficacia del RI temprano.

En la etapa 2C, 3 evaluadoras (2 Psicólogas y 1 Licenciada) *repite*ron la calificación de los ítems de precisión, coherencia y relevancia con el 30% de las aplicaciones, usando videograbaciones. Los objetivos de la segunda evaluación fueron: estudiar fuentes de error en la medición, posiblemente asociados a sesgos de objetividad de la primera evaluadora, y ampliar la perspectiva de interpretación de los puntajes de eficacia del RI, para proponer al IMERIN como herramienta confiable para el diagnóstico de habilidades inferenciales tempranas.

Estudio 3. La etapa 3A se realizó una aplicación para probar el funcionamiento del cuestionario sobre prácticas de socialización de la cognición. El contenido fue evaluado por los 12 participantes en la prueba, ellos dieron su concepto sobre la representatividad de las actividades de búsqueda de conocimiento deseable para el aprendizaje, las prácticas pedagógicas y las actividades comunicativas para la formación del pensamiento. En la etapa 3B los contenidos de la encuesta fueron evaluados por dos Estadísticos con experiencia en procesos de encuestas, ellos hicieron sugerencias a las escalas para elegir respuestas. En la etapa 3C se aplicó la encuesta a los 81 participantes y se analizaron los resultados mediante estadísticas descriptivas. Finalmente se describieron tendencias en los hábitos de comunicación y en las actividades esporádicas de búsqueda de conocimiento socialmente aceptado.

Creación del IMERIN-v1

Estudio 1: Diseño y prueba de contenidos para un instrumento de medición de atributos del RI

La construcción de las tareas del IMERIN-v1 se fundamentó en la necesidad de facilitar la evaluación del desarrollo de habilidades inferenciales de los niños de Preescolar. Se propuso para lograr este fin, evaluar la eficacia en el RI, mediante el uso de actividades de dominios STEM. Por esta razón los formatos de los contenidos fueron imágenes, figuras 3D y gráficas, que presentaban elementos de categorías, propiedades distintivas, funciones y sistemas de relaciones propios de cada dominio (Lakin y Gambrell, 2012).

La etapa *1a*, empezó con la Revisión Sistemática de la Literatura sobre estudios empíricos de la capacidad de RI en niños de Preescolar y concluyó con la conformación de una muestra de

120 actividades, con contenidos en 4 dominios de la Ciencias (Astronomía, Biología, Física y Estadística), en Tecnología y en Psicología. En síntesis, los principales objetivos de la etapa, fueron:

- a. Determinar la viabilidad del estudio a partir del análisis de las aplicaciones de prueba en las aulas. Esto fue necesario dado que no hay antecedentes de estudios similares con comunidades de estrato socioeconómico bajo, en Colombia ni en otros países latinoamericanos.
- b. Analizar contenidos e ítems que permiten evaluar la plausibilidad de generalizaciones a partir de categorización de propiedades, elementos y casos característicos de las relaciones razonadas.

Para este proceso se consideró que desde el enfoque de la Educación STEM se sugiere asegurar, además de la representatividad de las representaciones gráficas, semánticas y físicas, la perspectiva de igualdad de géneros. Lo anterior es especialmente importante en la creación de materiales con usos pedagógicos, o en indagaciones en las que los niños y niñas experimenten con las leyes o principios científicos. Dado que existe evidencia de creencias estereotipadas sobre el desempeño de las niñas en las actividades STEM (OECD, 2015).

Para efectos de precisar el aporte del trabajo metodológico del estudio 1 al desarrollo de la investigación se resumen los principios de construcción de las tareas de generalización, así: el conocimiento conceptual se generaliza a un conjunto de evidencia contenida en un sistema semántico (Kemp y Jern, 2014). Por lo tanto, los contenidos del instrumento, en especial el conjunto de evidencias de las relaciones objeto de razonamiento, deberá presentar características diferenciales tales como: diversidad- homogeneidad, casos típicos – atípicos, elementos conocidos- desconocidos. En la Tabla 2 se resumen principios conceptuales que orientaron la evaluación de desempeños en la prueba de contenidos del IMERIN- v1.

Tabla 2.

Criterios de comparación entre respuesta y principios del IMERIN-v1.

Dimensión de la Inferencia	Definiciones en los dominios de STEM: Física y Biología Respuestas esperadas según definición	Autores siglo XXI
Identificar regularidades espacio-temporales de los eventos y relaciones funcionales	Hacer analogías sobre la base de la similitud estructural o relacional Emparejar elementos a partir de relacionales parentales, de condiciones de hábitat y descendencia, aunque prevalezcan las coincidencias perceptivas y de apariencia Extender el conocimiento de similitudes a nuevas relaciones	Noyes & Christie, 2016 Slone & Johnson, 2018 Waismeyer & Meltzoff, 2017 Walker, Lombrozo, Legare & Gopnik, 2014
Categorizar nuevos miembros, propiedades y funciones de los sistemas	Experimentar con variedad de usos para los elementos que componen la tarea Favorecer manipulación de elementos o imágenes útiles, en vez de imágenes de dibujos sin sentido en su cotidianidad Extender el conocimiento de cualidades a nuevos componentes del sistema	Lane & Shafto, 2017 McCormack, Bramley, Frosch, Patrick, & Lagnado, 2016 Rhodes & Liebenson, 2015
Generalizar evidencia de casos conocidos a nuevos casos	Diferenciar entre las propiedades comunes, funcionales que identifican a los componentes del sistema o categoría, de las características situacionales, temporales o accidentales Asociar elementos involucrados en relaciones causales simples Extender el conocimiento de relaciones a nuevos sistemas	Gelman, Leslie, Was & Koch, 2015 Lawson, 2017 Sutherland & Cimpian, 2017

Posteriormente, en *la etapa 1b*, los criterios del diseño de actividades de RI fueron evaluados por expertos en los 6 dominios. Se eligieron 18 actividades de RI para crear contenidos de prueba en 3 dimensiones. Estas estaban conformadas por 72 de los 120 ítems iniciales. Los contenidos seleccionados constituyeron la versión de prueba del instrumento, IMERIN-v1. En la etapa 1c se determinó y probó el funcionamiento de la organización general (macroestructura) de las situaciones del RI, a partir del análisis de los indicadores de plausibilidad, precisión y coherencia definidos en los antecedentes, en las modalidades declarativa, procedimental y estratégica. De esa forma se crearon ítems para cada uno de los componentes de la respuesta explícita: intervención de un sistema de relaciones, elección de evidencia y adición de información en la verbalización de la conclusión, en las tres modalidades de interés.

En la etapa 1d se hizo la aplicación de prueba de 12, de las 18 actividades evaluadas por expertos en dominios STEM. Las profesoras de Preescolar valoraron la suficiencia de dos escalas de puntuación en el instrumento: una calificó la cantidad de aciertos por modalidad y la otra la coincidencia entre respuesta e indicadores de plausibilidad en cada ítem. Como resultado de la aplicación se construyó la versión inicial de la escala de medición de atributos de la eficacia en el

RI. En la Tabla 3 se presentan los referentes para la especificación de los criterios de comparación de coincidencia de la respuesta del niño con los principios implicados en las situaciones de RI del IMERIN-v1. Esta versión se cambió en las posteriores adecuaciones del instrumento a las modalidades de respuesta verbalizada.

Tabla 3.

Especificaciones de procesos involucrados en las tareas del IMERIN-v1

Modalidad inferencia	Respuesta Explícita	Atributo Eficacia	Expresión adicional	Tareas
Generalización	Categorización	Plausibilidad	Juicio	18
Anticipación Predicción	Intervención	Precisión	Exposición de ideas Recuento	6
Decisión	Elección	Coherencia	Justificación	6
Comparación Discriminación	Terminar proceso Completar secuencia	Relevancia	Explicación Supuesto	6

Participantes

-En la construcción y adaptación de tareas participaron: 8 expertos en docencia universitaria e investigación en áreas STEM, ellos son 1 mujer y 7 hombres, 5 Doctores en alguna Ciencia o en Ingeniería; y 2 Licenciadas en Preescolar, con más de 5 años de experiencia en el ejercicio docente y la enseñanza de las Ciencias en primera infancia. Ellos fueron convocados por su idoneidad en el campo de estudio y por sus antecedentes en investigación o en docencia en las áreas STEM. Su participación fue voluntaria, por interés en la innovación en la Educación del siglo XXI.

-La evaluación del funcionamiento de las actividades y de la representatividad de los indicadores de puntuación, fue realizada con la participación de 3 profesoras, Licenciadas en Educación Infantil, con experiencia en enseñanza de las ciencias. Las profesoras fueron invitadas a participar teniendo en cuenta su amplia experiencia en el uso de herramientas para la evaluación en Educación Preescolar, o en la atención a la Primera Infancia en situación de vulnerabilidad socioeconómica. Su participación consistió en aplicar algunas actividades del IMERIN-v1 y en aportar conceptos, valoraciones y criterios para la calificación de los ítems. Además, la participación consistió en ser entrevistadas. En la entrevista focalizada se indagó, en profundidad, las prácticas de búsqueda de evidencia y conocimiento sobre regularidades estadísticas cotidianas en los entornos académicos y familiares de sus estudiantes.

-En la prueba piloto (de funcionamiento) de los materiales participaron 31 menores, 16 niños y 15 niñas, con edades entre 4 y 4.5 años ($M= 4.2$), de los cursos del nivel Jardín de un Colegio Distrital de Bogotá, ubicado en una localidad de estratos socioeconómicos 1 y 2. Algunos de los participantes eran extranjeros y pertenecían a familias desplazadas por la violencia. Para su vinculación se obtuvo el consentimiento informado de los padres, o madres y el asentimiento informado de los niños.

Diseño

La evaluación de prueba de los contenidos del IMERIN-v1, se realizó por segmentos. De manera que cada una de las 4 evaluadoras, probó el funcionamiento de 3 o 4 muestras de las actividades de cada modalidad, de forma independiente, en 2 grupos de niños. La aplicación se hizo individualmente, en el aula. Los niños participantes se asignaron al azar a cada grupo. Al final todos los niños participaron en al menos 3 de las actividades. La etapa se realizó durante el primer semestre de 2019.

Procedimiento

Las actividades de familiarización y las prueba de los materiales y situaciones fueron una especie de laboratorio de experiencias de RI, realizadas durante la jornada escolar. De manera que las actividades se integraron a los procesos de aula y las respuestas de los niños ocurrieron de manera espontánea. Se siguió el diseño en la aplicación, procurando que la presencia de la investigadora se hiciera habitual para los niños, así como los experimentos intuitivos con los materiales y contenidos del IMERIN-v1.

Instrumentos y recursos de la aplicación de prueba del IMERIN-v1

Se diseñaron 4 clases de contenidos, cada clase tenía entre 4 y 6 tareas, además se creó un conjunto de 90 ítems para generalización, se evaluaron 36 de ellos. Las Figuras 2 a 5 contienen imágenes de las actividades de categorización que fueron usadas en la fase de familiarización, en las aulas, con el grupo de niños y el acompañamiento de la profesora directora del curso. En las Figuras 6 y 7 se muestran ejemplos de los ítems de generalización del IMERIN- v1, que también fueron aplicados en el aula. Ejemplos de estos ítems se presentan en los Anexos A hasta E.

Resultados de la aplicación de prueba

El resultado de la aplicación se analizó cualitativamente y mediante el porcentaje de coincidencia en las puntuaciones entre observadoras independientes, éste estuvo por encima de 90 % en las tres comparaciones. Se usó también el coeficiente de correlación Spearman, ($Rho=0.8$). Posteriormente, el análisis se centró en el significado de las coincidencias y discrepancias entre las evaluadoras, respecto a la pertinencia de los contenidos, y entre las conclusiones de los niños y los indicadores de puntuación.

Puntuación de la subescala generalización. La calificación de la dimensión se obtuvo sumando los puntos de cada ítem, el puntaje total de la fase de generalización se expresó como el porcentaje de ítems que recibieron calificación. En términos cualitativos, se evaluó el porcentaje como un desempeño de nivel bajo, medio o alto. El objetivo último fue transformar el conjunto de calificaciones a una medida que pudiera ser interpretada como la cuantificación de la plausibilidad de la generalización expresada.

Evaluación de desempeño en las actividades. La evaluación de las inferencias en cada ítem se realizó comparando la coincidencia de respuesta con los indicadores de eficacia. Posteriormente se asignó un punto a cada coincidencia y a partir del porcentaje de ítems que recibieron calificación se estimó un desempeño de nivel bajo, medio o alto. El principal hallazgo del proceso metodológico fue el acuerdo en que los contenidos y materiales tienen las cualidades *pertinencia y objetividad*, es decir, que permiten que las respuestas del razonador sean contrastadas frente a la explicación conceptual y a las definiciones propias del dominio, las cuales deben coincidir con hechos verificables. Estos logros se concretaron en los indicadores de eficacia en la inferencia, que serían usados en las demás versiones del IMERIN.

Una vez se habían realizado los ajustes necesarios a esta versión del IMERIN se presentarían los contenidos a los jueces expertos, en marzo de 2020. Tras la crisis sanitaria y el cierre de instituciones educativas fue necesario replantear los contenidos y modalidades de la inferencia explícita a evaluar. Por esta razón se creó un nuevo instrumento.

Actividad de categorización: el objetivo fue estudiar la variedad de respuestas explícitas de los niños de 4 años, usando imágenes de elementos de categorías familiares y de procesos, probablemente desconocidos para ellos.

Indicación: el conjunto de imágenes de la derecha se completa con una de las imágenes del grupo de la izquierda. El niño deberá elegir la imagen que crea que muestra el elemento faltante.

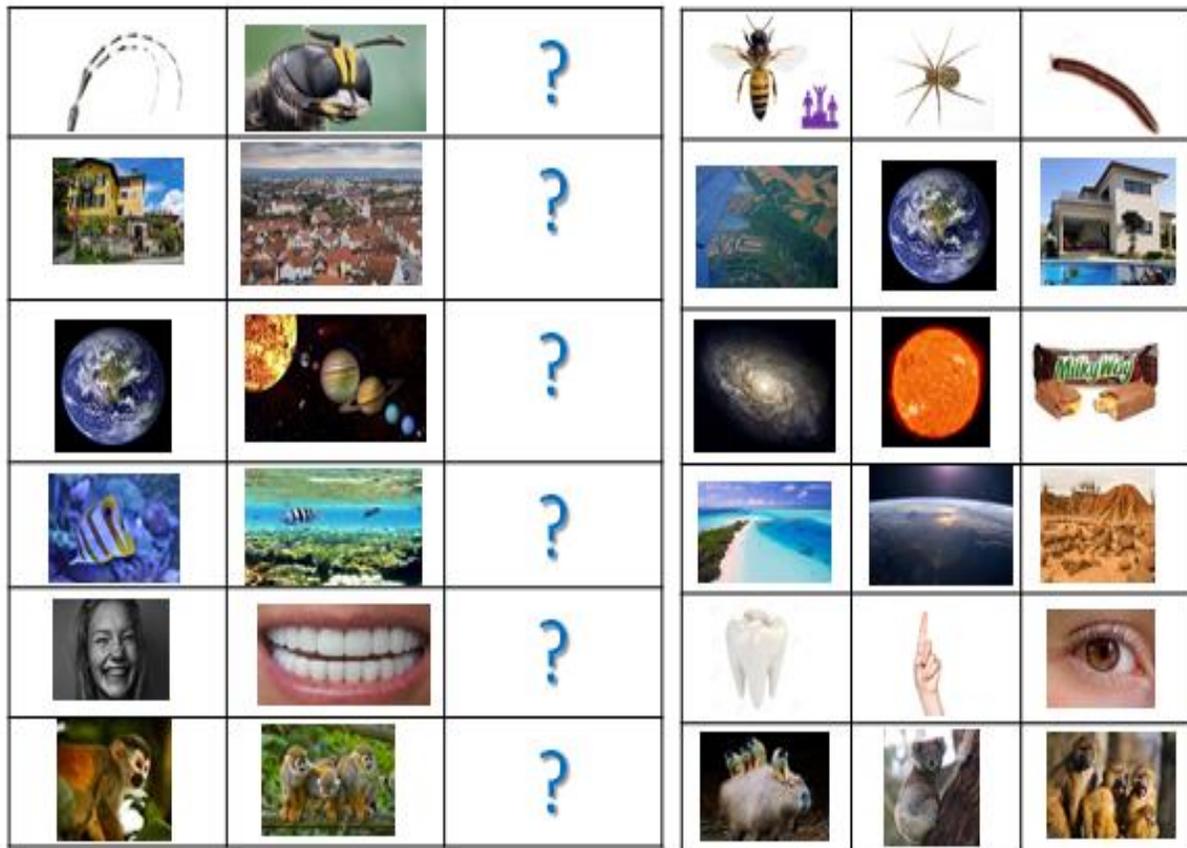


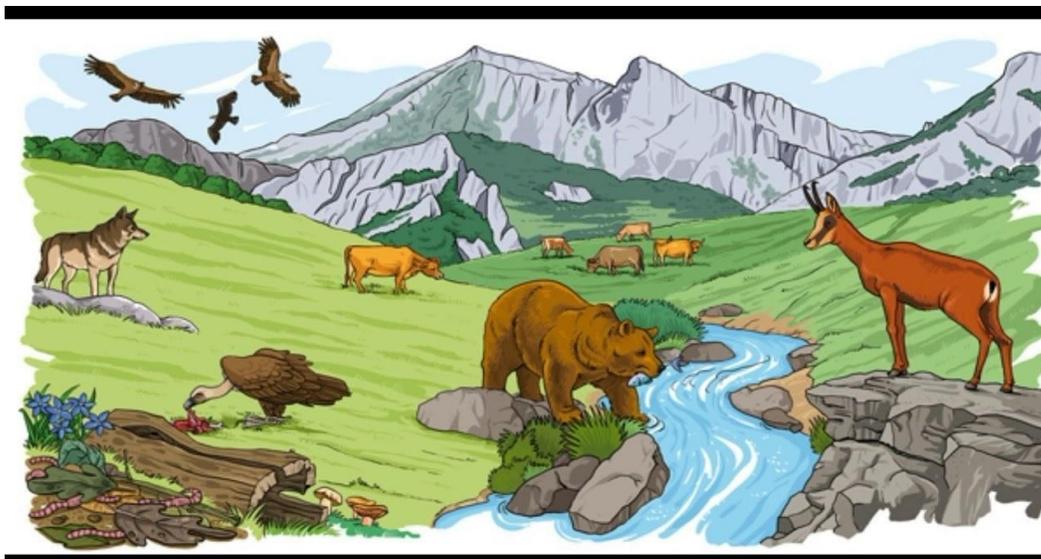
Figura 2. Actividades de categorización del IMERIN –v1 fase de familiarización.



Figura 3. Actividades de categorización IMERIN-v1 fase de familiarización

Actividad de categorización: el objetivo fue estudiar la variedad de respuestas explícitas en la discriminación de propiedades, usando imágenes de elementos de categorías familiares, entre las cuales hay algunas que no pertenecen a la categoría.

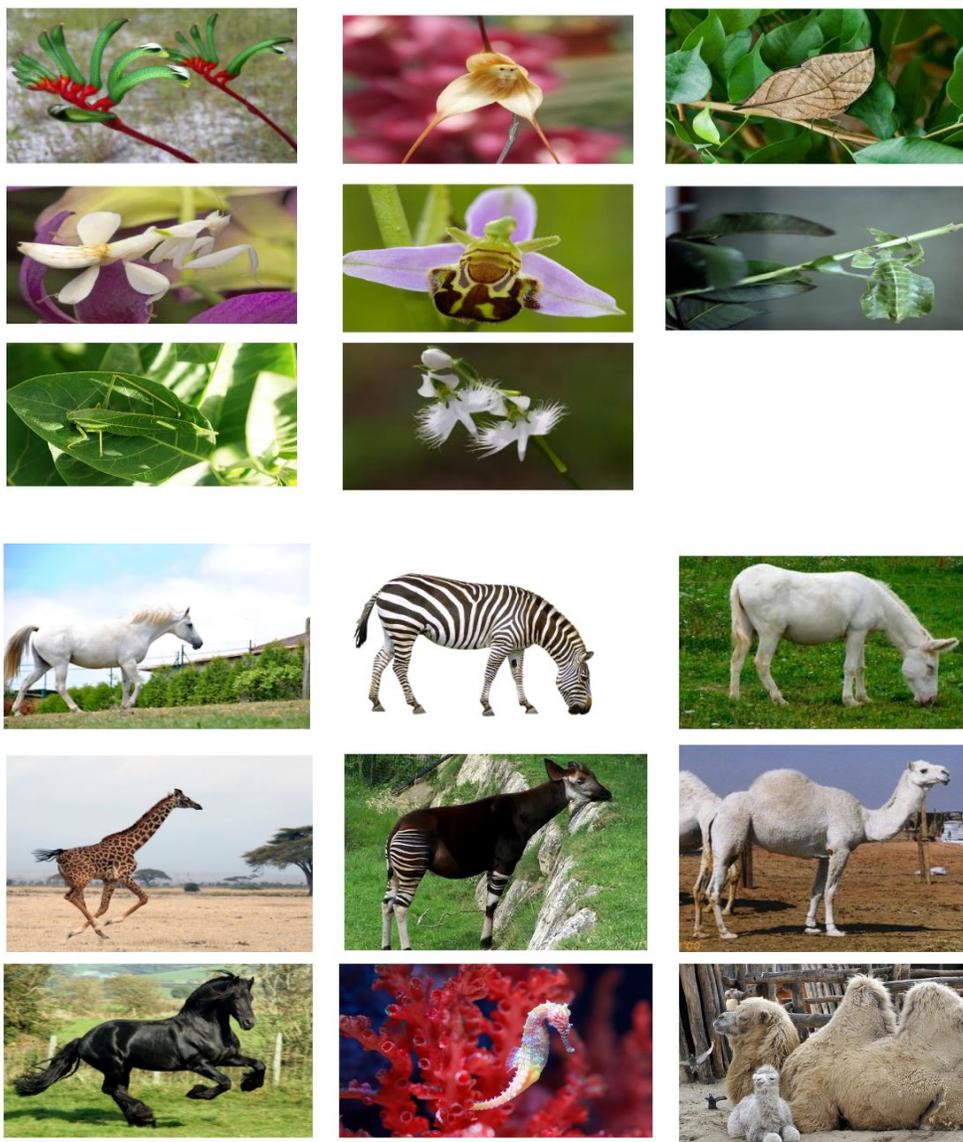
Indicación: entre el conjunto de imágenes hay unas distintas que no pertenecen a la misma categoría. El niño deberá armar el conjunto, descartando la imagen que crea que muestra el elemento diferente.



Actividad de familiarización con conceptos de dominio específico: el objetivo fue estudiar la variedad de explicaciones que expresan los niños durante tareas de emparejamiento.

Indicación: entre el conjunto de imágenes hay unas que coinciden exactamente unas con otras, son fragmentos de un proceso, o parejas de una escena, o dos elementos distintivos de las propiedades distintivas de categorías. El niño deberá armar la pareja, explicando si coinciden exactamente, o tienen algo diferente y porque es esa la que completa la imagen.

Figuras 4 y 5. Actividades de categorización por emparejamiento.



El objetivo de las actividades de generalización a partir de categorías diversas, fue verificar que los participantes estaban en capacidad de inferir casos representativos de las propiedades distintivas de la clase y a la vez, casos representativos de la excepción. La inferencia plausible implica evaluar evidencia implícita del contexto, probablemente usando conocimiento conceptual de las categorías, del cual no se presenta información en los contenidos.

Figuras 6 y 7. Imágenes de las secuencias usadas para las actividades de generalización de propiedades de categorización, del IMERIN -v1

Estudio 1- etapa 2. Diseño de nuevos contenidos del IMERIN-P

El cambio en la metodología obedeció a la necesidad de adaptar las estrategias de evaluación de plausibilidad de las inferencias y los criterios de medición de indicadores de eficacia de las conclusiones a las nuevas condiciones de aplicación. El uso de tecnologías y medios de comunicación remota se eligió bajo el supuesto de que la aplicación a distancia del IMERIN-P permitiría obtener datos y observaciones confiables. Se planteó la expectativa de que nuevas estrategias de indagación permitirían obtener nueva evidencia de las características de la inferencia inductiva, que probablemente no se habían observado en los estudios anteriores (Sternberg & Prezt, 2005).

Participantes

Durante toda la etapa participaron: 9 niños de 4, 5 y 6 años; 3 (de las 9) madres; 3 Licenciadas en Preescolar, expertas en enseñanza de las ciencias en primera infancia (2 de ellas madres de niños participantes

Participación en etapa 2a. Las profesoras evaluaron los textos de los indicadores de las actividades de RI y emitieron su concepto sobre la objetividad y suficiencia de estos. Además, dieron ejemplos de otras formas explícitas que podría tener la conclusión de los niños (hipótesis, explicación y predicción)

Participación en etapa 2b. los niños realizaron las actividades de prueba durante una videollamada, sus madres los acompañaron, pero no participaron en la aplicación.

Participación en etapa 2c. inicialmente, las 3 profesoras elaboraron listas de palabras para la etapa de Redes Semánticas, luego las madres usaron estas listas para elaborar otras palabras con significados semejantes. Finalmente, las profesoras participaron en la evaluación de la representatividad de los ítems de los instrumentos teniendo en cuenta la etapa de Redes Semánticas.

Instrumentos y Procedimientos

Una vez se determinó que el proceso de prueba y aplicación para validación del IMERIN-P era viable, dado que se disponía del apoyo de las profesoras y directivos de las instituciones convocadas en 2019, además de otros recursos (TIC, por ejemplo), se avanzó en la actualización de la literatura sobre los principios conceptuales y metodológicos necesarios para el diseño de ítems e indicadores de los dos nuevos instrumentos. Cumplidos los objetivos iniciales se integraron al marco teórico dichas especificaciones, las cuales se enfocan en las singularidades de las respuestas

explícitas verbalizadas y están suscritas en contextos de covariación y causalidad probabilística, en dominios de las Ciencias, la Tecnología, la Estadística y la Psicología. (Anexos H y K).

Tabla 4.

Características de la evidencia en relación con el dominio de la actividad del IMERIN-P.

Generalización		
Casos en los que es evidente la covariación, o la no covariación en las propiedades, relaciones o elementos del sistema de causalidad. Situaciones que pueden ser cotidianas y presentan regularidades estadísticas, consideradas con alta probabilidad de haber sido conocidas por los niños en sus contextos.		
Hipótesis – evidencia indirecta	Explicación – evidencia directa	Predicción – evidencia probabilística
Ciencias 1 Representan la variabilidad y la transición entre estados de categorías implicadas en la generalización previa, incluso contradiciendo expectativas	Ciencias 2 Cambios en dominio de los resultados, que pueden ser explicados como efecto de propiedades no observadas, de intenciones humanas, por ejemplo.	Ciencias 3 Cambios en relaciones de causalidad, en las que se puede intervenir para alterar, a futuro, el efecto observado
Estadística 4 Características de las muestras, frecuencias y datos estadísticos que tienen potencial de estar relacionados con un resultado conocido que no se observa en los contenidos	Estadística 5 Casos de excepción a la regla de relación, no generalizable aunque se observen correlaciones frecuentes entre eventos que parecen similares, pero son de diferentes dominios	Estadística 6 Categorías recientemente aprendidas se usan como causas suficientes del efecto razonado, pero estas tienen probabilidad condicional de ocurrir juntas a futuro
Tecnología 7 El mecanismo que explican la relación de causalidad es invariante, aunque cambien la información de representatividad de los datos - muestra (producto)	Tecnología 8 Una intervención posibilita la observación del efecto, pero este es causado por propiedades no obvias de los objetos sobre los que se interviene	Tecnología 9 Cambios en las condiciones de los mecanismos de relación que no causan cambios en los principios, pero tienen probabilidad de generar variaciones del resultado a futuro
Psicología Todas las actividades presentan evidencia de la agencia en el sistema de relaciones razonado, dado que ésta es una propiedad generalizable, no incidental. Razonar sobre la agencia humana implica consciencia del potencial de generar conocimiento verificable.		

Etapa 2a. se optimizaron los indicadores para evaluar las modalidades de la conclusión (hipótesis, explicación y predicción), a partir de los resultados de la actualización conceptual.

Etapa 2b. Se hizo una *aplicación de prueba a distancia* para estudiar el funcionamiento de los contenidos, con la participación de 9 niños. Al final del IMERIN-P se encontraba una escala de apreciación de la seguridad, o confianza, que sintió el participante frente a sus conclusiones. Estaba compuesta por 3 imágenes: una tenía la mano cerrada y el pulgar arriba y se indicó que servía para mostrar mucha confianza en haber hecho “buenas conclusiones”; otra era la imagen de una mano cerrada, con todos los dedos juntos, se debía usar para indicar que no sentía ni confianza ni desconfianza de sus aciertos o fracasos en la respuesta; y la mano con el pulgar abajo se podía usar como señal de poca seguridad frente a las conclusiones, o de desconfianza en haber dado respuestas acertadas. En la versión final del instrumento este ítem no aparece, la razón se explica más adelante.

En la Tabla 4 se presenta una síntesis de las características de la evidencia en relación con el dominio del contenido en la actividad del IMERIN-P. La Tabla 5 presenta la Matriz de especificaciones que orientaron la creación de nuevos indicadores de eficacia, estos se usaron para evaluar el desempeño en el IMERIN-P.

Etapa 2c. El ajuste de indicadores se realizó mediante la técnica de Redes semánticas (Vera Noriega et al, 2005) y la evaluación de representatividad de las imágenes en los contenidos, con la participación de profesoras y madres. Mediante ésta técnica se crearon 2 conjuntos de palabras. El primero de palabras con significados similares a los conceptos involucrados en los indicadores de eficacia y el segundo, con palabras referidas a los significados de comportamientos infantiles relacionadas con dichos conceptos. La elección de esta técnica se justificó por su utilidad en la generación de indicadores de *precisión semántica*, los autores citados se refieren al proceso de creación de redes semánticas como “*la etapa de los significados culturales*”. (Anexo F)

Tabla 5.

Matriz de especificaciones para evaluación del desempeño en el IMERIN-P.

Indicadores criterios de medición a validar	Procesos subyacentes al RI eficaz	Ejemplos situaciones de evaluación	Ejemplos ítems modalidad respuesta verbal
Plausibilidad de verificación de hechos y afirmaciones que el niño incluye en la conclusión	Comprensión de evidencia de las características y propiedades de los objetos, artefactos y agentes	Información y evidencias de:	Contenidos Representaciones visuales de fenómenos naturales, físicos, químicos y sociales con múltiples cambios
		Eficacia causal en sistemas de relaciones entre variables continuas: tiempo, espacio, luz, energía y agencia humana	
Precisión: Coincidencia entre las nociones y el conocimiento conceptual usado para explicar relaciones de causalidad y los principios – leyes del dominio	Intención de hacer inferencias válidas. Uso del conocimiento de regularidades y cambio en interacciones espacio-temporales, motivaciones y otras transiciones las relaciones	Conocimiento de Covariación y frecuencia de eventos cotidianos y de cambios en el entorno	Videos para recreación de probabilidades de eventos fisicoquímicos y electromagnéticos, en condiciones inciertas o parcialmente observadas
		Regularidades y datos estadísticos aprendidos de la observación de hechos cotidianos	Evidencia directa de covariación y contingencia en procesos biológicos vitales y no vitales, o estados de la materia, entre otros
Coherencia entre la representación mental del niño y la evidencia directa –imagen		Capacidad de agencia propia – intervenir las relaciones de un sistema de causalidad para cambiarlo críticamente	Escenificación de situaciones de comunicación, interacción y comportamientos que dan cuenta de cambios en la relación inferida, o en estados mentales de los agentes
Relevancia: usar info. y conocimiento implícito (creencia) para controlar incertidumbre frente a evidencia indirecta	Uso de teorías ingenuas		

El procedimiento para construir las redes fue el siguiente:

1. Se solicitó a cada uno de los expertos disciplinares y profesores que asesoraron la creación de contenidos de las primeras tareas de RI, que elaboraran un listado de palabras, de uso popular, referentes a los contenidos y en especial a las reglas, o principios de las relaciones implícitas en los ítems.

2. Posteriormente, dichos listados fueron usados por cada una de las profesoras participantes como referencia para proponer otras palabras, referidas a los comportamientos de los niños, que creyeran asociadas al significado de las primeras. Se les solicitó considerar las palabras que podrían ser usadas frecuentemente, o que resultaran más familiares para los niños y las personas del común.

3. Finalmente se repitió el ejercicio con 3 de las madres de los niños que participaron en la prueba de funcionamiento del IMERIN-P, usando las palabras que listaron las profesoras. Se les pidió que incluyeran las palabras que creían que sus hijos expresarían al referirse a las palabras de la lista elaborada por las profesoras.

Conclusiones sobre la confiabilidad de los indicadores de eficacia en la inferencia. Los funcionamientos de los indicadores de eficacia corroboraron el supuesto de que, dado un contenido que presenta evidencia de relaciones de causalidad en dominios específicos, solo un conjunto de inferencias y ciertas conclusiones sobre esas relaciones pueden ser plausibles. Dichas conclusiones están determinadas en el marco explicativo de la causalidad proporcionado por el dominio específico (Wellman et al, 2015). Este marco delimita la razonabilidad (justificación) de las afirmaciones seleccionadas como premisas a representar en el sistema de relaciones probabilísticas, que son la base de los contenidos. En la Tabla 6 se presenta una síntesis de los criterios de evaluación usados para verificar la coincidencia de la conclusión verbalizada y los indicadores de acierto en el ítem. IMERIN-P

Tabla 6.

Criterios de evaluación de la coincidencia de la conclusión verbalizada y los indicadores de acierto en el ítem. IMERIN-P

Modalidades	Hipótesis	Explicación	Predicción
Dimensiones			
Precisión de propiedades definitorias de categorías	Coincidencia entre las nuevas categorías o propiedades verbalizadas y las que tienen probabilidad de ser elementos del sistema de causalidad	Especificación del mecanismo que hace posible la relación, como covariación o no – covariación, correlación, antecedentes espacio-temporales y otros que coinciden	Exposición de condiciones previas que aumentan la probabilidad de que las categorías estén implicadas en futuros resultados o sistemas de causalidad
Coherencia	Coincidencia entre el conocimiento semántico explícito sobre nueva evidencia y los principios abstractos que fundamentan las evidencias anteriores – fase 1	Atribución de causalidad a propiedades, intervenciones o mecanismos causales que coinciden con evidencia directa y son verificables	Correspondencia entre la anticipación de hechos y relaciones que pueden repetirse dado el patrón de la nueva evidencia
Relevancia Integración de información relevante	Referencia a la provisionalidad de la conclusión, o a que no sea posible hacer una conclusión definitiva por falta de certeza	La nueva información agregada es análoga, comparable y útil para reducir la incertidumbre y aclarar o descubrir principios causales implícitos en el evento	Proponer nuevos datos comparables con la evidencia indirecta de las condiciones que favorecen la repetición de eventos a futuro

Conclusiones no representativas de alguna dimensión de la eficacia en el RI

Centrar la conclusión en la similitud entre propiedades de la apariencia, exteriores o evidentes, o en la asociación de eventos (Goswami 2010).	Consideración de asociaciones sobrenaturales, extraordinarias o fantásticas entre los eventos; sin posibilidad de verificación (Lane 2016)	Expresar creencias sesgadas en relación con las categorías y relaciones de causalidad en los eventos y sistemas razonados. Integrar información que aumente la incertidumbre (Erickson et al, 2010)
---	--	---

Especificación de criterios e indicadores de evaluación de la conclusión verbalizada.

Los criterios de evaluación de atributos del RI fueron definidos a partir de la revisión de antecedentes, se confirmó su pertinencia en el estudio 1. Los indicadores de plausibilidad, precisión, coherencia y relevancia en la inferencia se desprenden de los criterios de evaluación y se usaron para la asignación de puntaje en la dimensión. Para la prueba piloto se eligieron los que

resultaron más favorables para comparar la respuesta. Estos enunciados son descripciones, *semánticamente precisas*, de las categorías conceptuales y de las reglas de asociación implicadas en las relaciones subyacentes a la evidencia. Además, los indicadores contienen datos, e información estadística verificable por comparación con las conceptualizaciones, o explicaciones de la teoría objetiva (del dominio).

Las conclusiones verbalizadas son evidencia de los desempeños en las diferentes modalidades de la inferencia explícita. Estos también pueden evaluarse a partir de comparaciones con indicadores de plausibilidad. Por ejemplo se compararon las siguientes respuestas: a) afirmaciones que describen la intervención para crear cambios en las situaciones; b) descripciones de propiedades o relaciones que hacen explícita su conclusión sobre causalidad; y c) enunciaciones, identificación de agentes, o de componentes de los sistemas de relaciones, que intervienen en la covariación, incluso de aquellos que no hacen parte del sistema ni de las condiciones que intervienen en el resultado (Gopnik et al, 2001; Sobel et al, 2017; Sternberg y Prezt, 2005).

Escala de confianza en las conclusiones. Aunque todos los participantes respondieron a la pregunta por el sentimiento de confianza en sus conclusiones, los resultados se invalidaron por intervención de las personas que acompañaron la aplicación. Dado que las respuestas mostraron no ser representativas de la conducta que pretendía observarse, ésta escala se eliminó del instrumento.

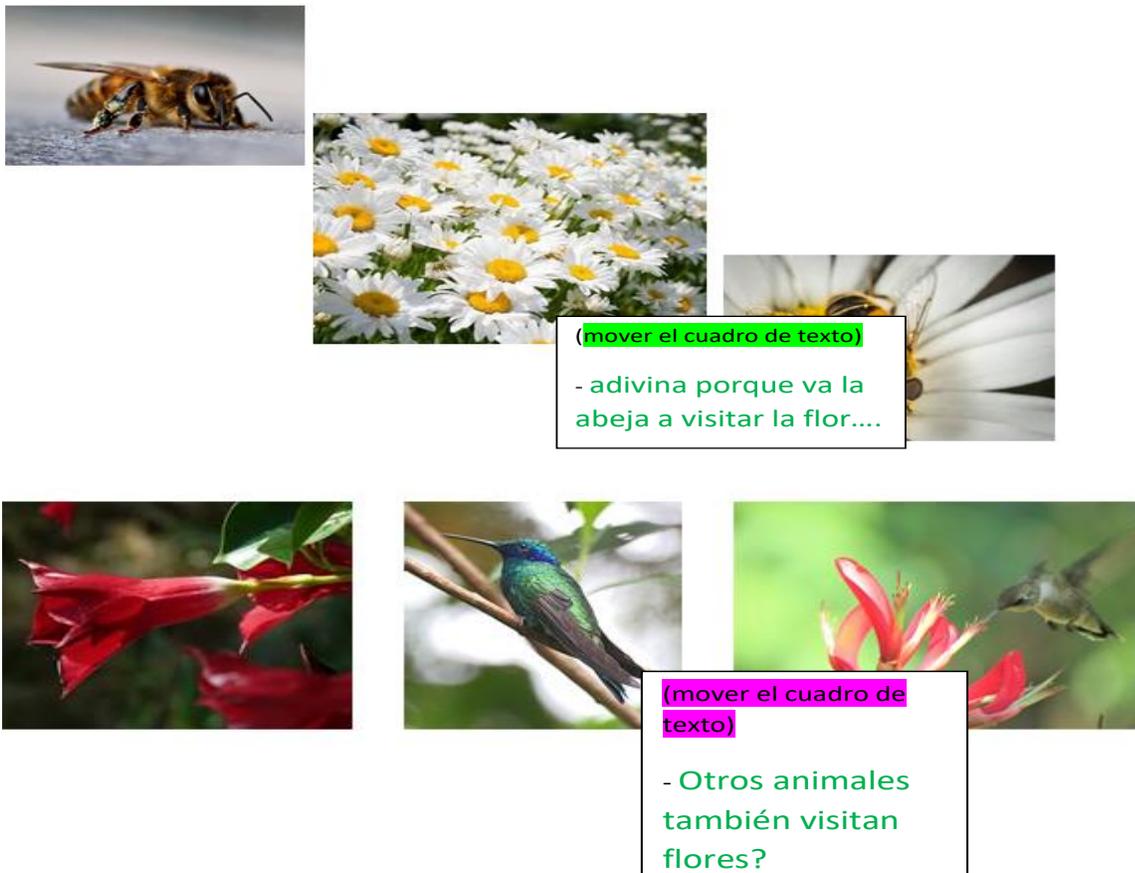
Contenidos elegidos para fase de generalización del IMERIN -P. Los contenidos elegidos a partir de los resultados de la prueba de contenidos para generalización (Figuras 8 y 9) se usarían en los ítems del IMERIN-P, presentado posteriormente al juicio de validación. La elección de los ítems y el formato final de las actividades de RI fue asesorada por expertos en enseñanza de las disciplinas STEM, en todos los niveles de educación. Además, ellos participaron en la evaluación de la precisión semántica de los contenidos gráficos. Las profesoras directoras de curso (Jardín) evaluaron la pertinencia de las instrucciones y orientaciones verbales que se darían al participante durante la actividad.

El criterio fundamental para la elección de los contenidos para presentar a juicio fue la precisión semántica de éstos. Los contenidos semánticamente precisos, permiten indagar el razonamiento sensible al contexto (Slousky et al, 2015). Puesto que presentan evidencia directa de regularidades, relaciones causales, correlación y covariación entre eventos, en dominios específicos, sobre los cuales razonan los niños en la cotidianidad. En ese orden de ideas, el alcance

y posibilidad de verificación de las conclusiones en un dominio específico, sustentan la conexión lógica entre las diversas clases de contenidos de la prueba (Schulz, 2012).

En general, los nuevos ítems presentaban evidencia directa e indirecta de cambios en la probabilidad de relación y ocurrencia de eventos en formatos visuales exclusivamente. Se usaron videos y fotografías de escenarios reales en los que covarían elementos de la situación y se observan cambios temporales y en los patrones de dependencia-independencia entre elementos. Las diferencias entre los contenidos de las mismas dimensiones se dan en función de las instrucciones y las modalidades explícitas de las conclusiones a observar. Estas se introdujeron con el fin de evaluar consistencia de la inferencia (más adelante sería un índice de confiabilidad de la prueba).

Ejemplos de tareas de generalización: el objetivo fue generalizar la relación, biológicamente determinada, entre la actividad de alimentación de insectos y aves, y la luz solar, o la iluminación del día.



Figuras 8 y 9. Ejemplos de los ítems de generalización del IMERIN- P.

Contenidos fase de prueba. Específicamente, cada ítem está conformado por un conjunto de imágenes, una indicación, solicitud o pregunta focal que restringe la respuesta y dos enunciados que son indicadores de las generalizaciones plausibles y de las inferencias eficaces, según la dimensión evaluada. Cualquiera de los dos indicadores puede ser usado como criterio para comparar la coincidencia en la respuesta y decidir si se asigna la calificación de 0 o 1 en el ítem. En la Figura 10 se presenta un ejemplo de las actividades que componían el IMERIN-P, juzgado para validación de contenidos.

Fases de las actividades de RI. La estructura de las actividades del IMERIN se diseñó a partir de las recomendaciones de Simon y Lea (1974). Específicamente, respecto a la importancia de diferenciar entre las estructuras de situaciones de inducción y la organización de tareas de solución de problemas. De manera que, las actividades de RI se deben realizar en al menos 2 fases con diferentes objetivos interrelacionados. En ese orden de ideas, cada actividad de RI se compone de 4 ítems, que se aplican en 2 fases. Es importante mencionar, que la estructura inductiva de las actividades para medición de la eficacia en la inferencia, permite el control de la dificultad y de la demanda de recursos cognitivos de los ítems. Esto favorece la evaluación justa, en términos del desarrollo de Funciones Ejecutivas y del lenguaje, esperado para la edad de los participantes (Colberg et al, 1982).

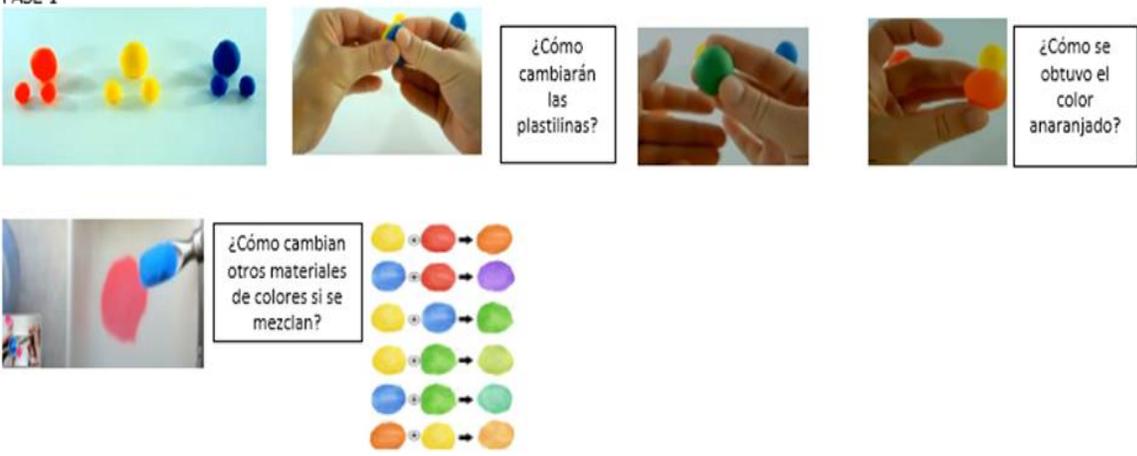
El procesamiento de información en la fase inicial del RI consiste en traducir evidencias de las reglas de relación, implícitas en los contenidos, en representaciones conceptuales del dominio específico de la actividad (Leighton y Sternber, 2004). Los niños pueden inferir conocimiento generalizable, porque a la evidencia presentada en esta fase, subyace un patrón o regla abstracta que regula las relaciones sobre las que se razona (Schultz y Gopnik, 2004).

En la primera parte de esta fase se presentan las categorías y sus denominaciones. El objetivo fue familiarizar a los niños con las evidencias diversas y las propiedades invisibles, en caso de que fueran desconocidas para ellos. Por esta razón se usan imágenes y videos que hacen énfasis en los procesos, o en las propiedades de los objetos, artefactos y agentes distintivos de las categorías implicadas en la relación. Se usó contenido visual con el fin de favorecer que el participante se oriente por sus teorías intuitivas (Shafto et al., 2008). En la segunda parte de fase 1, la de generalización de las reglas de relación, los contenidos presentan un nuevo caso de la relación, o información de un nuevo resultado de la relación. Algunos de los contenidos no tienen probabilidad de ser evidencia de un caso que abarque la generalidad. Dependiendo del dominio y

la modalidad de la inferencia observada, las preguntas se focalizan en la naturaleza del contenido, o en las consecuencias de la relación representada.

Durante la fase de prueba se genera evidencia de que los niños están en capacidad de coordinar múltiples inferencias. Permitiendo además, confirmar si al RI subyace la comprensión de la evidencia, teorías y creencias relevantes, o algún conocimiento del dominio específico al que pertenecen los contenidos. De hecho, algunos expertos en el área sugieren que, si la actividad cognitiva se limita a un solo tipo de inferencia, la posibilidad de observar y estimar atributos del razonamiento se restringe (Ricco, 2015). El objetivo de la fase es evaluar la habilidad de razonar sobre la probabilidad de cambios de estado y de transiciones sistemáticas en las relaciones, de forma eficaz. Por lo tanto, los 3 ítems de la fase 2 se usan para estimar precisión, coherencia y relevancia de la conclusión. Los contenidos de esta fase presentan evidencia diversa, directa e indirecta y casos de covariaciones, o variabilidad en las condiciones de la relación generalizada en la fase 1.

TAREA No 1: TRASFORMACIÓN DE LA MATERIA POR INTERVENCIÓN HUMANA
FASE 1



¿Cómo cambiarán las plastilinas?

¿Cómo se obtuvo el color anaranjado?

¿Cómo cambian otros materiales de colores si se mezclan?

FASE 2



¿Cuáles colores mezclaron para hacer este osito?

¿Se puede transformar el osito en otra figurita de color amarillo?

Figura 10. Ejemplo de una actividad del IMERIN-P.

Algunos ítems no son representativos del tipo de relación inferida previamente, e incluso contienen casos de excepción a la generalización hecha en la fase anterior. En general, los ítems de prueba se caracterizan por presentar contenidos que apoyan diferentes hipótesis igualmente plausibles. Esto implica el uso de información adicional a la contenida en la evidencia, para reducir o controlar la incertidumbre y mejorar la probabilidad de la conclusión.

Para estimar la plausibilidad de la generalización y la eficacia en el RI se debe asignar puntuación en cada ítem. El criterio para asignar puntaje es la coincidencia entre la conclusión y alguno de los indicadores de acierto en el ítem. De esa forma se obtuvieron las siguientes estimaciones: a) frecuencia de aciertos en cada dominio y modalidad de la inferencia evaluada; b) porcentaje de aciertos en cada dominio y modalidad; c) proporción de conclusiones que expresaron conocimiento preciso; d) proporción de respuestas que se basaron en comparación, o integración coherente, de las nuevas evidencias con las conclusiones anteriores; y e) la proporción de conclusiones que aportaron información relevante para la conclusión. En la Figura 11, se presenta un ejemplo de las actividades del IMERIN-P.

Versión del IMERIN para juicio de validación de contenidos. El producto final del estudio 2b es el IMERIN-P, el cual quedó conformado por: a) 9 series de imágenes, cada una compuesta por series de entre 3 y 6 fotografías (presentación de Microsoft PowerPoint), una de las cuales tiene, además, un video (en InShot); b) un protocolo de aplicación que contiene las instrucciones, consignas y preguntas de cada ítem, fueron 36 ítems en total; y c) un manual de calificación de las respuestas, que contiene las instrucciones para el uso de los indicadores y para el registro del puntaje de eficacia, por parte del evaluador. Además, entre el material se encuentra espacio para el registro de las observaciones adicionales que resulten de importancia en el análisis de comportamientos o factores que puedan influir el desempeño de los niños en la actividad.

Estudio 2: Indagación de validez y confiabilidad del IMERIN y sus usos propuestos

Con el objetivo de generar evidencia de la utilidad científica del instrumento de medición para la caracterización de habilidad inferencial en dominios STEM, en esta etapa de la investigación se llevaron a cabo 3 procesos de evaluación de validez y confiabilidad de la instrumentalización. Esto implicó generar evidencia empírica a partir de la estimación de eficacia en la inferencia inductiva, de los criterios de interpretación del puntaje, del procedimiento y de la documentación de la medición.

Fase 1 familiarización: generalización de hundimiento como resultado de meter piedras y de empujar esponjas al agua.



Se presenta el Video 1: se observa una piedra negra hundiéndose y una esponja amarilla que se empuja entre el agua, después de que ésta se hunde, se exprime y se deja flotando

¿Qué puede ocurrir cuando se mete una piedra blanca y se empuja una esponja rosada entre agua?

Indicadores de plausibilidad en la generalización basada en similitudes

-Plantear la expectativa de que cualquier piedra que entre al agua se hundirá (podría postular que la causa es el peso de esta) y que todas las esponjas se hundan si son empujadas

- Suponer que la intervención del agente que pone en contacto los elementos causa el evento hundirse, por lo que este efecto volverá a ocurrir si se repite la intervención con elementos semejantes

- Referirse al video 1 para concluir que la piedra blanca y la esponja rosada se hundirán como ya se observó.

El criterio para pasar a la siguiente fase es que se verifique 1 de los 3 indicadores de la generalización

Fase 2: Inferir la influencia de propiedades desconocidas, no obvias o invisibles de los nuevos elementos, proponiendo nuevas causas-inhedoras que evitan que se repita el efecto observado -la expectativa-



Video 2: muestra una piedra blanca hundiéndose y 2 piedras pómez, similares en apariencia a una esponja, estas no se hundan a pesar de que se empujan



Enunciado: Las piedras pómez son especiales porque flotan aunque son piedras...hay de varias formas y colores



Item 1. ¿Por qué estas no se hundan? Posterior a la respuesta se explica que los cuadrados son de piedra pómez

Indicadores de Precisión

Supone nuevas propiedades de los objetos, de acuerdo al cambio de color, las cuales podrían explicar que floten en vez de hundirse

Postula nuevas categorías, o propiedades del agua o del contenedor que podrían explicar el nuevo efecto

Item 2. ¿Por qué flotan las piedras pómez si son de piedra?

Indicadores de Coherencia

Plantea la hipótesis de que sean piedras "artificiales" fabricadas para flotar en el agua en vez de hundirse

Postula una nueva relación causal entre la denominación "pómez" y la inhibición del efecto hundirse en el agua, aunque se empuje

Item 3. ¿Para qué se habrán cambiado las esponjas por piedras pómez?

Indicadores de Relevancia

Hace hipótesis sobre la intención del agente (del video) de sorprender al observador con el efecto no esperado de "flotación"

Proponer que el cambio fue planeado con intención de demostrar diferentes propiedades de las piedras que parecen esponjas pero no lo son

El criterio para asignar un (1) punto en cada ítem es que se verifique 1 de los 2 indicadores de la dimensión

Figura 11. Ejemplo de una actividad del IMERIN-P descrita en el Manual de aplicación. Nota: esta actividad forma parte de la versión final del instrumento.

Etapa 2a Validación de contenidos: Juicio de expertos en investigación del desarrollo cognitivo

La verificación de supuestos sobre el adecuado funcionamiento de las actividades de RI en la prueba piloto justificó avanzar al juicio de validación de contenidos. El objetivo del juicio de expertos fue obtener evidencia de la validez de los contenidos del IMERIN-P. La validación de contenidos consistió en evaluar la pertinencia, suficiencia, objetividad y adecuada redacción de los componentes de 9 actividades de RI en dominios STEM. El resultado confirmó que las actividades seleccionadas por los jueces, permiten estimar la capacidad de generalización y medir atributos de la eficacia en la inferencia, a partir de la evaluación de las verbalizaciones de hipótesis, explicaciones y predicciones. El juicio de Psicólogos expertos se realizó durante el semestre I de 2021.

De acuerdo con los objetivos metodológicos, el juicio tuvo dos énfasis: a) evaluación de la coherencia entre la conceptualización de los mecanismos subyacentes a la eficacia en el RI y el diseño metodológico para su estimación; y b) usar las categorías de juicio pertinencia, suficiencia, objetividad y adecuada redacción, como criterio para calificar los ítems de cada dimensión de la eficacia y sus indicadores, estos últimos usados para la puntuación de la conclusión.

Como parte de la evaluación de validez de contenidos del IMERIN-P se solicitó a los jueces considerar la correspondencia entre los principios teóricos que definen el RI eficaz, las modalidades de la inferencia estudiadas y las estrategias llevadas a cabo para instrumentalizar el proceso de medición de atributos de la eficacia. Para establecer dicha correspondencia es necesario juzgar la coherencia conceptual de los criterios para la estimación de eficacia y de los componentes tanto del IMERIN, como del procedimiento mismo de medición. A partir de los conceptos y calificaciones de los jueces se estableció la pertinencia de la interpretación de indicadores de eficacia, de las escalas de medida y del uso sugerido para los índices de habilidad obtenidos mediante los puntajes del instrumento.

Adicionalmente, se solicitó a los jueces hacer sugerencias al procedimiento, o a las herramientas de evaluación según los criterios de validez externa que consideraran pertinentes. Puesto que se propone el uso del IMERIN como prueba diagnóstica de desarrollo temprano de habilidades inferenciales críticas en el aprendizaje de los dominios STEM (AERA, APA y NCME, 2014). De esta manera, las valoraciones de los jueces también aportaron elementos para analizar

factores que podrían influir en la varianza relevante, o esperada, en los puntajes y otras fuentes que podrían originar varianza irrelevante, o errores de estimación.

Participantes

Los 4 jueces expertos en Psicología Cognitiva y de la Educación, son 2 mujeres y 2 hombres, psicólogos, doctores en Psicología o en Educación y están vinculados a la producción científica y académica de Colombia. Fueron convocados por su idoneidad en el campo de estudio y por sus antecedentes en la investigación empírica de la cognición en la infancia, en dominios específicos y en contextos educativos. Su participación fue voluntaria, por interés y aceptación de la invitación a conformar el grupo de jueces.

Recursos e instrumentos

Como resultado de los estudios anteriores se consolidó un conjunto de ítems para la evaluar plausibilidad de las generalizaciones y estimar atributos de la eficacia en la inferencia de los niños. A partir de estos contenidos se conformaron 9 actividades de RI en 3 dominios STEM, las cuales constituyen el IMERIN –P. Además de las actividades, los contenidos para la evaluación incluyeron: a) las especificaciones conceptuales y metodológicas de las modalidades de inferencia a evaluar (unidades de análisis), usadas para la construcción de las situaciones de RI; b) las presentaciones de los ítems aplicados en la prueba piloto de los contenidos; y c) los protocolos y formatos de calificación y registro de observaciones, para diligenciar durante el procedimiento de aplicación (manual). Los jueces recibieron el instrumento y demás documentos en formato digital. Las categorías propuestas para fundamentar el juicio fueron definidas de la siguiente forma:

Pertinencia: esta condición determina si las actividades son aptas para obtener evidencia directa y representativa de los comportamientos distintivos de las inferencias eficaces en la inferencia inductiva, en cada una de las formas explícitas de ésta. Como propiedad de los contenidos, la pertinencia depende de la relación conceptual entre los principios que definen el atributo medido y los indicadores propuestos como criterio para puntuar el ítem. La pertinencia puede ser juzgada en términos de conveniencia, si las dimensiones y los ítems propuestos coinciden exactamente con las medidas de eficacia, que han sido validadas por la investigación empírica antecedente, la coincidencia significa que los ítems propuestos son convenientes para la medición requerida.

Suficiencia: por medio de esta categoría se juzga si los contenidos de las situaciones de RI son aquellos indispensables para guiar la inferencia de relaciones y principios que dan plausibilidad

a la conclusión. Si son suficientes, los ítems contienen una completa y óptima variedad de evidencia sobre los elementos y sistemas objeto de razonamiento, de manera que permitan la observación de las 3 modalidades de la conclusión verbalizada. Además, la suficiencia puede ser atribuida a la posibilidad que ofrecen los contenidos de transformar las respuestas del niño (observaciones) en datos, asignándoles valores numéricos a las unidades de análisis o a las conclusiones, con el fin de crear escalas y subescalas de calificación o de medición de los atributos.

Objetividad: los indicadores son objetivos si permiten que las respuestas del razonador sean contrastadas frente a la explicación conceptual y a las definiciones propias del dominio, las cuales deben coincidir con hechos verificables. En la aplicación del instrumento la objetividad se garantiza si el criterio de puntuación es aplicado únicamente a la conclusión. Dado que las conclusiones plausibles han sido predeterminadas, los indicadores son independientes de la valoración subjetiva del evaluador que los use. Es decir, la objetividad es una condición que aplica a los indicadores de acierto en la conclusión. Si son objetivos, los indicadores de precisión, coherencia y relevancia pueden ser usados para la caracterización de atributos de la cognición en dominios STEM, por parte de investigadores, educadores y observadores competentes.

Adecuación de forma y redacción: se refiere a la valoración de representatividad y calidad de las imágenes, al nivel de exactitud de las instrucciones, preguntas e indicaciones dadas a los participantes. Además, a la claridad en los textos descriptivos de la situación, cuya lectura corresponde al evaluador y buscan facilitar la interpretación de la actividad por parte del participante. En general, los contenidos adecuadamente presentados y redactados permiten a niñas y niños comprender las tareas, preguntas y modalidad de expresión de las respuestas solicitadas.

Para la clasificación de los ítems en cada nivel de la categoría de juicio se usó una escala de calificación con puntajes entre 0 y 3, los criterios generales para la puntuación fueron los siguientes: si el ítem no cumple con las condiciones para calificar en la categoría, se puntuó 0; si el ítem cumple con las condiciones para calificar en el nivel bajo, se puntuó 1; si el ítem cumple con las condiciones para calificar en el nivel medio, se puntuó 2; y si el ítem cumple todas las condiciones para calificar en el nivel alto, se puntuó 3. Por ejemplo, para la dimensión “pertinencia” en cada actividad se especificaron las siguientes calificaciones:

- 0: Si el ítem no es pertinente, si es inútil para la medición de la dimensión
- 1: Si el ítem tiene un nivel bajo de pertinencia - utilidad para la medición de la dimensión
- 2: Si el ítem tiene un nivel medio de pertinencia - utilidad para la medición de la dimensión
- 3: Si el ítem es muy pertinente, tiene un nivel alto de utilidad para medir la dimensión

Procedimiento

Una vez recopilados los documentos elaborados por los jueces, la información sobre su juicio individual se procesó mediante 3 estrategias. Primero se analizó la evaluación de coherencia conceptual de la definición operacional de la inferencia, en sus modalidades explícitas (hipótesis, explicación y predicción) y la coherencia con la metodología propuesta para evaluar su plausibilidad (comparación con indicadores). Luego se sumaron las calificaciones de cada ítem de la actividad, para establecer cuáles fueron las mejores puntuadas y cuáles requerían ser ajustadas según sus sugerencias. Finalmente se estableció si los conceptos fueron favorables sobre la utilidad de la instrumentalización para estimar eficacia de la inferencia, es decir, si se juzgó confiable el uso sugerido para el puntaje y su interpretación.

Los análisis estadísticos del resultado cuantitativo del juicio se usaron para justificar decisiones metodológicas, principalmente, sobre elección de actividades para conformar la versión del instrumento a ser aplicado para completar la validación. Además, los conceptos y sugerencias se incorporaron al plan de análisis de resultados, de acuerdo con el objetivo de describir propiedades psicométricas del IMERIN.

Conceptos de los jueces respecto a la fundamentación teórica del IMERIN. Como parte de la validación de contenidos del instrumento se evaluó positivamente la conceptualización de los atributos de la eficacia y de los criterios para evaluar la plausibilidad de las conclusiones generalizantes. La evaluación de los jueces para el conjunto de ítems de generalización significa que es pertinente considerar lo como el ítem de filtro, es decir, éste debe superarse para seguir la aplicación de la fase de prueba. El uso propuesto para esta clase de ítems se justifica puesto que, la capacidad de hacer inferencias generalizantes, plausibles, es precursora de la habilidad inferencial inductiva (Goswami, 2010).

Los conceptos de los jueces, respecto a la fundamentación teórica de la estrategia de instrumentalización propuesta para la estimación de eficacia en la inferencia, fueron favorables. Así mismo, se confirmó acuerdo en los juicios sobre la pertinencia y suficiencia de las dimensiones precisión, coherencia y relevancia como atributos de la eficacia. Este conjunto de resultados indica que los ítems son confiables para la medición propuesta. En conjunto, sus evaluaciones y juicios cualitativos se interpretaron con el propósito de mejorar el procedimiento de medición y verificar que se consideraron factores críticos que podrían influir en la varianza relevante, y la no relevante, de los puntajes del instrumento (AERA, APA y NCME 2014; Elosua, 2003). De igual manera, sus

observaciones sobre el documento descriptivo del instrumento fueron positivas, confirmando los supuestos sobre la adecuada sistematización, orden y diseño. Es importante resaltar que los elementos referidos están asociados a la validez ecológica del estudio.

2b Aplicación del IMERIN: validación de la propuesta de interpretación de puntajes

El objetivo general de la aplicación de validación fue generar evidencia empírica que sustente la propuesta de usar los resultados del IMERIN con fines científicos. Para lograr este objetivo se aplicó el instrumento y se estimaron niveles de la eficacia en el RI. Por lo tanto, las puntuaciones son útiles para clasificar los desempeños (generalizaciones, explicaciones y predicciones) y para caracterizar diferencias individuales en los perfiles de RI. Siguiendo el objetivo y a partir del resultado del juicio de expertos, se anticipó que los puntajes individuales mostrarían diferentes patrones en las magnitudes en cada dimensión de la eficacia. A partir de la medición se obtuvo un conjunto de puntajes que permite describir tendencias en el desarrollo de habilidades inferenciales inductivas de un grupo de niños colombianos, con edades entre 4 y 6 años.

A partir del juicio de expertos se obtuvo evidencia de la confiabilidad del procedimiento para la estimación requerida, sea que ésta se lleve a cabo presencial, o a distancia. Sin embargo, se consideró la necesidad de realizar una prueba de funcionalidad del instrumento en 2 tipos de aplicación. Puesto que, las condiciones de obtención de la evidencia cambiaron durante el periodo de crisis social ocasionada por la Pandemia y el paro nacional de 2021. Por esta razón se aplicó con diferentes grupos de participantes, para determinar, mediante herramientas de análisis estadístico de las puntuaciones, la independencia de la medida y de su procedimiento de estimación, de las particularidades de la situación de observación del desempeño.

Participantes

La aplicación para validación del IMERIN se realizó individualmente, con ciento ocho (108) participantes en total, 48 niñas y 60 niños, con edades entre 4 y 6 años ($M=4,5$ años, $D.E.=0,62$). Todos ellos estaban matriculados en Jardín o Transición, durante el I de 2021, en tres tipos de instituciones prestadoras de servicios educativos en el nivel Preescolar, así: en 2 Instituciones Distritales 17 niñas y 31 niños; en una institución privada 19 niñas y 12 niños; y en una de convenio (una modalidad que administra el Distrito y financia parcialmente una

organización de carácter privado) 12 niñas y 17 niños, la Tabla 7 muestra la distribución de los participantes por género, grupo o nivel y tipo de institución en los que estaban inscritos.

Las instituciones fueron seleccionadas a partir de muestreo por conveniencia y los participantes fueron contactados a través de las profesoras directoras de cada curso, ellas mediaron la convocatoria y permitieron acceder a encuentros virtuales con las madres y acudientes, para informarlos y obtener sus datos de contacto. Las madres y padres fueron contactados vía telefónica y los interesados recibieron los formularios (de Google Docs.) de inscripción, estos presentaban los objetivos de la investigación y la descripción de las formas de participación (de los menores de edad y de ellos), y registrar su consentimiento informado. También se solicitó asentimiento informado a las niñas y los niños, después de explicarles las actividades en las que se les invitó a participar.

Durante la aplicación 2 niños participantes abandonaron y 3 niñas y 2 niños pasaron a modo de actividad de aprendizaje por cumplir el criterio para suspender la fase de prueba (completar 3 ítems seguidos sin obtener puntuación). Posteriormente, por dificultades de salud asociadas al Covid-19 o por falta de medios para realizar la videollamada, una niña y un niño superaron los 6 años, 1 mes y 3 participantes más se retiraron de la institución. Aunque con este grupo de 5 niños se realizó la actividad, cuando estuvieron en disposición, sus calificaciones no se incluyeron en la muestra para el estudio de propiedades psicométricas del IMERIN. De manera que los puntajes considerados para los análisis estadísticos corresponden a 96 aplicaciones, los otros 12 desempeños se analizaron en función de la caracterización de diferencias de desempeño en los dominios.

Diseño

La aplicación se efectuó siguiendo un diseño 2x3x2, con las modalidades de aplicación (presencial o no presencial) como variable distintiva entre los tres tipos de institución (distrital, privada o de convenio) y diferenciando el orden de aplicación según el género del participante y el nivel de Preescolar (Jardín y Transición) en el que se encontraba.

Tabla 7.

Composición de la muestra de participantes en la aplicación de validación del IMERIN.

Institución	Grupo	Presencial		No - Presencial		Total Grupos		Total Institución
		Niñas	Niños	Niñas	Niños	Niñas	Niños	
Distrital	Jardín	0	0	4	12	4	12	48
	Transición	0	0	13	19	13	19	
Privada	Jardín	8	6	1	0	9	6	31
	Transición	7	6	1	2	8	8	
Convenio	Jardín	10	11	2	4	12	15	29
	Transición	0	0	0	2	0	2	
Total		25	21	20	38	45	59	108

Instrumento y materiales

El IMERIN consta de: a) 6 actividades, cada una se compone de entre 3 y 6 fotografías (presentación de Microsoft PowerPoint) y una tiene además un video (en InShot); b) un protocolo de aplicación que contiene las instrucciones, consignas y preguntas de cada uno de los 24 ítems; y c) un manual de calificación, que contiene las instrucciones para el uso de los indicadores y para el registro del puntaje de eficacia. Entre el material se encuentra espacio para el registro de observaciones adicionales que resulten de importancia en el análisis del comportamiento y de los factores que puedan influir el desempeño de los niños en la actividad. Las Figuras 12 a 17 explican contenidos de 3 actividades del IMERIN validado.

Actividad Ciencias – Explicación

El objetivo de la actividad es evaluar el RI sobre variaciones en los efectos de relaciones causales, a partir de evidencias que facilitan:

- a) generalizar el efecto - "camuflarse" o esconderse, como resultado de la interacción entre propiedades de la apariencia y de la conducta de los insectos.
- b) Inferir variaciones y una excepción en los resultados del camuflaje, sin relación con necesidades vitales, aclarando expectativas sobre nuevas causas de la semejanza en la apariencia

Contexto: La conducta "camuflaje" de algunos insectos es una adaptación biológica, por lo tanto, tiene un vínculo causal con la sobrevivencia de la especie. Las imitaciones de apariencia que hacen las personas cuando se disfrazan son motivadas por deseos y no representan relaciones de causalidad en el dominio de la Biología. Aunque cambiar de apariencia, probablemente, favorece un efecto similar de ocultamiento de la persona, éste cambio no está relacionado con el camuflaje en los insectos.

Instrucciones para la aplicación - fase de familiarización:

- 1 se presentan las imágenes de tres insectos camuflados entre plantas



2 refiriéndose a las tres imágenes de insectos camuflados, preguntar:

- **Cuál es el ser vivo que observamos en la fotografía?**
- 3 se señala la presencia del insecto y se informan los nombres

4 Pregunta para evaluar la generalización:

- **¿Por qué estos insectos se quedan encima de las flores o de las hojas?**

Indicadores de generalización de la relación apariencia - conducta – efecto:

- Explicar que la ubicación de los insectos se relaciona con búsqueda de escondites adecuados para realizar conductas como cazar, comer, dormir, o cualquiera que se relacione con causar efectos en otros animales
- Proponer que la semejanza en la apariencia del insecto y de las hojas o la flor, causa que estos permanezcan entre las plantas, para lograr algún objetivo asociado a la presencia de otros insectos, animales o personas, incluso por defensa
- Afirmar que la apariencia de los insectos les permite esconderse entre las plantas y flores para no ser vistos fácilmente por otros animales

Las conclusiones plausibles son evidencia de que se infieren nuevos casos en los que resulta relevante la similitud para camuflarse y causar un efecto que favorece la adaptación del insecto

Fase de prueba ítem de precisión



Indicaciones

5. Esta Mantis busca un lugar para camuflarse y esperar a que pasen insectos para atraparlos...

- ¿Cómo debe ser el sitio para que la Mantis no se vea fácilmente?

Indicadores de Precisión en la conclusión

- Proponer características del sitio similares a la apariencia de la Mantis, como color verde, o ser parte de una planta o de una flor
- Describir espacios que permitan a la Mantis acercarse, cubrirse o posarse sobre hojas, para que otros insectos no la vean

Fase de prueba ítem de coherencia



¿Es bueno el lugar que encontró la Mantis, Por qué si / no?

Indicador de Coherencia

Afirmar que es difícil ver la Mantis y este es el resultado de estar bien camuflada, por lo que es probable que consiga atrapar insectos que se acerquen sin verla

Explicar que el color de las flores podría servir para que otros insectos vieran a la Mantis, a pesar de que se parece a las hojas

Fase de prueba ítem de relevancia



¿Porque estas personas se disfrazaron de insectos?

Indicadores de Relevancia

Atribuir el cambio de apariencia a la intención de participar en situaciones sociales que requieren lucir como un insecto

Proponer el gusto por los insectos, la voluntad, o la diversión, para explicar el cambio temporal de apariencia

Actividad Estadística- Explicación

Objetivo: evaluar el RI sobre regularidades en fenómenos de dominios diferentes, que tienen alta probabilidad de ocurrencia y covariación. Sin embargo, no son casos representativos de relaciones causales. Implica inferir excepciones en la regla generalizada, a partir de la evaluación de evidencia sobre casos en los que la relación no tiene probabilidad, e incluso casos en los que se invierte la probabilidad. Además, los contenidos podrían asociarse con creencias sesgadas sobre las actividades nocturnas de algunas categorías de animales. Por lo que, la revisión de dichas creencias, podría ser necesaria para mejorar la plausibilidad de las conclusiones.

Contexto: con frecuencia los cambios en la actividad humana y de otros seres vivos se relacionan con variaciones de la intensidad de luz, o con la oscuridad. Esta relación es probabilística, por lo tanto, no se explica por relaciones de causalidad. Las condiciones que explican estas covariaciones son comunes, u ordinarias en el dominio social, o de la Biología, por lo que no es necesario proponer algún tipo de causas extraordinarias.

Fase de familiarización: **en estas fotografías, ¿es de día o de noche?**



Objetivo de Ítem de generalización: generalizar vínculos entre las variaciones regulares en los fenómenos astronómicos y cambios regulares en los comportamientos de animales y humanos

¿Qué actividades realizan los seres vivos en la noche?, ¿Por qué no hacen esas actividades de día?

Indicadores de generalización plausible

- Explicar que los seres vivos no tienen actividad nocturna, porque cuando es de noche duermen
- Relacionar la visualización del sol y la luna con cambios en las actividades de vigilia – sueño de los animales
- Afirmar que todos los seres vivos cambian de actividad en la noche, porque está oscuro, o porque no hay luz

El criterio para pasar a la siguiente fase es que se verifique 1 de 3 indicadores de la generalización plausible

Objetivo fase de prueba: Inferir nuevos casos de covariación regular en conductas de animales, como resultado de cambios de iluminación en el día y la noche, y casos de excepción a la regularidad en conductas de animales y de humanos

Introducción: **Las flores del cactus crecen de día y de noche, si reciben luz del sol o de la luna, pero solo en la noche las visitan murciélagos...**



Fase de prueba – ítem de precisión



¿Por qué el murciélago solo visita la flor en la noche?

Indicadores de Precisión

Plantear vínculo entre la oscuridad de la noche, o la poca luz de la luna y la búsqueda de alimento de los murciélagos

Aclarar que en la noche ocurren cambios, como poca iluminación, y que éstos se relacionan con actividades de alimentación de los murciélagos

Fase de prueba ítem de coherencia



¿Las abejas y los colibrís también se alimentan de las flores en la noche?, ¿Por qué si / por qué no?

Indicadores de Coherencia

Aclarar que colibrís o abejas solo se alimentan en el día, cuando hay condiciones asociadas a la permanencia del sol, como temperatura o calor

Describir algunas condiciones climáticas del día que favorecen la búsqueda de flores o la alimentación, y que no se presentan en la noche

Fase de prueba ítem de relevancia

Introducción: Las personas podemos observar estrellas en la noche, si no hay nubes en el cielo



¿Cuáles actividades podemos hacer las personas únicamente de día? ¿Por qué?

Indicadores de Relevancia

Describir actividades condicionadas a estados del clima, o a la presencia del sol, como salir a pasear, a broncearse, o a tomar el sol

Nombrar actividades preferidas, o hábitos que se realizan en la cotidianidad, dependiendo de si es de día, si hay luz solar, o de condiciones climáticas; si hace calor por ejemplo

Actividad Ciencias – Predicción

Objetivo: evaluar el RI sobre evidencias de cambios en las conductas de autocuidado y probabilidades de revertir daños en los dientes. Anticipando condiciones que pueden intervenir para recuperar la salud dental. Usando inferencias de nuevos casos de la relación, para predecir el tipo de conocimiento experto que podría requerirse para prevenir, el mismo tipo de daño, en diferentes clases de dientes mamíferos.

Contexto: las condiciones adecuadas de cuidado, higiene y alimentación pueden ser factores de protección de la salud. Estas disminuyen riesgos de cambios dañinos para los dientes. El uso del conocimiento y tecnología especializados aumentan la probabilidad de revertir cambios, en las estructuras dentales, aparentemente irreversibles.

Fase familiarización objetivo: identificación de cambios en los dientes asociados con el desarrollo humano



Indicación: Los niños cambian de dientes... los dientes que tienen a los 5- 6 años, se caen, luego crecen unos nuevos que no vuelven a caerse si se cuidan muy bien

Objetivo item de generalización: generalizar la relación entre cambios en los dientes, conductas de autocuidado y prevención de daños

¿Qué deberías hacer para que tus nuevos dientes se mantengan sanos toda la vida y no se caigan?



Indicadores de plausibilidad en la generalización

- Anunciar que deberá mantener hábitos de higiene y de cuidado de la salud dental para evitar la caída de los nuevos dientes
- Anticipar que podrían ocurrir cambios naturales en los dientes, como crecimiento, o envejecimiento, o daños por descuido, que tal vez se podrían evitar realizando alguna conducta de prevención.
- Mencionar conductas que debe evitar, porque causaran daños, o accidentes y perdida de dientes

Objetivo fase de prueba: inferir nuevos casos de cambio de los dientes, correlacionados con repeticiones de conductas, cambios en higiene, o nuevas conductas de cuidado, incluso en diversas clases de dientes

Ítem de precisión

Introducción: Una persona que comía chocolates para sentirse feliz y se lavaba la boca pocas veces, perdió algunos dientes; por esto fue al Odontólogo y él los reemplazó con dientes artificiales, llamados prótesis dentales...



Después de esto, ¿esa persona volverá a perder dientes? Por qué si / por qué no?

Indicador de Precisión

Proponer que se mejorará el cuidado y se evitarían conductas similares de descuido en alimentación e higiene dental.

Suponer que se repetirán conductas que causaron daños, por lo que es probable que ocurran nuevas pérdidas, o que podrían ocurrir accidentes que dañen los dientes o las prótesis.

Fase de prueba Ítem de coherencia



Los Odontólogos podrían seguir ayudando a esta persona con sus dientes?

Indicador de Coherencia

- Suponer que esta persona acudirá de nuevo al Odontólogo para aprender a prevenir nuevos problemas, o por ayuda para cuidar las prótesis y los demás dientes que están sanos
- Afirmar que nuevos daños en ambas clases de dientes, ocasionados por consumo de chocolate, u otros alimentos, podrían ser reparados por la intervención de expertos

Fase de prueba Ítem de relevancia



- Si tuvieras un gato, deberás hacer algo para que sus dientes se mantengan sanos?

Indicadores de Relevancia

Anticipar que deberá aprender sobre la salud de los dientes de los gatos, o que podría acudir al Veterinario para que le enseñe cómo cuidarlos

Afirmar que evitará darle comidas con azúcar, o chocolates y buscará ayuda de expertos en mantener la higiene y cuidado de los dientes del gato

Figuras 12 a 17. Tres (de 6) actividades del IMERIN, versión validada por jueces.

Procedimiento

La aplicación fue individual y siguió el protocolo establecido, el orden y la duración fueron similares en las dos modalidades de ésta. Cada actividad se realizó en tres fases, según el diseño validado, el siguiente fue el orden de aplicación:

1. *Familiarización*. Para iniciar se presentaron imágenes de video o fotografías de elementos que distinguen las categorías, o casos distintivos de las relaciones implicadas en la situación objeto de razonamiento. El objetivo de esta fase es facilitar el recuerdo de conocimientos y representaciones semánticas de los contenidos. Por esta razón, se mencionaron las denominaciones de los eventos o de las propiedades, luego se hizo una pregunta para confirmar que los contenidos fueron identificados o recordados.

Se esperaba que el participante incluyera en la respuesta el nombre y las clases de objetos, artefactos y agentes presentes en las situaciones. Si no ocurría la respuesta esperada, se repetían las explicaciones y preguntas sobre las imágenes, esta vez dando un ejemplo de alguna situación, de la cotidianidad, en la que es probable que los niños observen contenidos similares a los de las imágenes. Aunque se incluyó una segunda explicación para cada una de las situaciones de familiarización de las 6 actividades no fue necesario usarla, en ninguna de las aplicaciones.

2. *Fase de Generalización - Ítem de filtro*. Los contenidos en estos ítems son representativos de un caso diverso de las categorías implicadas en la fase de familiarización y muestran evidencia directa de un efecto, o de una regla o principio que regula la interacción entre los elementos del sistema presentado. La pregunta evalúa, según el caso, la disponibilidad de conocimiento generalizable, semántico o de causalidad. Dado que se planteó como ítem de filtro, no se repitieron las frases de ilustración del contenido en caso de que el participante no hiciera una conclusión plausible en la primera respuesta.

La puntuación de la conclusión (juicio generalizante) se asignó si ésta se encontraba incluida en alguno de los dos indicadores de calificación en cada ítem. A los participantes que no generalizaron adecuadamente la evidencia en el primer intento se les invitó a observar de nuevo la evidencia, indicando que había “algo” en lo que debían pensar antes de volver a responder, luego se repitió la pregunta. El 98% de los 108 participantes resolvió adecuadamente esta fase en el primer intento y continuaron a la fase 2 de la actividad.

3. *Fase de prueba*. Se aplicó siguiendo el protocolo y asignando la puntuación a cada ítem según los criterios establecidos. La puntuación dependió de las coincidencias (aproximadas), entre

las conclusiones y alguno de los 2 indicadores de acierto en cada dimensión evaluada. A continuación, se presenta una síntesis de dichos indicadores:

a) uso de conocimiento conceptual *preciso*, sobre categorías, mecanismos o principios que regulan relaciones de causalidad, patrones de dependencia e independencia y covariaciones sistemáticas en eventos, entre los que se incluye comportamientos intencionales de agentes.

b) delimitación *coherente* de las conclusiones, estableciendo la coincidencia entre estas y la nueva evidencia presentada. Implica la comparación de nuevas representaciones de propiedades distintivas de las categorías, con conocimientos sobre regularidades en las relaciones y con evidencia directa, o conocida.

c) adición de nueva información *relevante* para el control de la incertidumbre. Se supone que las teorías intuitivas y las creencias de los niños son la fuente de información útil en la elección de alguna de las hipótesis que tienen probabilidad de ser plausibles, dada la evidencia.

La aplicación no presencial se programó con las madres a través de mensajes, o llamadas en la aplicación WhatsApp y se realizó por medio de videollamada, usando las plataformas Zoom, Microsoft Teams o Google Meet, conforme a su elección. Previamente se había recopilado información sociodemográfica de las familias de los participantes, usando un formulario Anexo al documento digital del consentimiento informado.

En la modalidad presencial se requirió el control de posibles influencias ambientales en la atención, disposición y desempeño de los participantes. Por esta razón se adecuó un salón de la institución que ofreció condiciones para el distanciamiento físico, silencio, iluminación y ventilación. Además, se dispuso de un computador portátil para la presentación de las actividades que componen el IMERIN y muebles para ubicación de las personas, los materiales de la aplicación y los implementos contemplados en los protocolos de bioseguridad, para la higiene y desinfección previa y posterior a cada aplicación.

Siguiendo el esquema, presentado en la Tabla 7, se realizó la aplicación en ambas modalidades. La duración promedio en cada actividad fue 2.50 minutos, de manera que el tiempo total estuvo entre 20 y 25 minutos aproximadamente (las 6 actividades). Se había planeado, para no superar esta duración y evitar la fatiga de los participantes, que la aplicación podría dividirse en dos sesiones, también se acordó con las madres y las profesoras una alternativa para el acompañamiento en caso de tener que hacer otra sesión. Sin embargo, en ninguna de las ocasiones

que se preguntó si deseaban parar la actividad se respondió afirmativamente, ni reportaron sentirse cansados.

Es importante mencionar que la aplicación en el aula mejoró la oportunidad de obtener evidencia de validez ecológica del instrumento, además fue posible probar su funcionamiento en condiciones que podrían favorecer la réplica del estudio, en el aula y en diferentes contextos que podrían interesar a educadores o investigadores cuyo objetivo fuera realizar indagaciones del desarrollo cognitivo situadas.

2c. Acuerdo interevaluadoras: validez de criterio de las subescalas de medición y validez predictiva del resultado

La segunda evaluación del 30% de las aplicaciones ($N=26$), tuvo dos objetivos: el primero fue generar evidencia empírica de la validez de criterio de la medida, estudiando la consistencia de las puntuaciones entre observadoras independientes. El segundo fue analizar la confiabilidad en la interpretación de desempeños y significado de las coincidencias y discrepancias entre conclusiones e indicadores de puntuación. Los perfiles de habilidades esbozados por las evaluadoras, a partir de las asignaciones de niveles de eficacia, se analizaron con el interés adicional en generar evidencia de la validez predictiva del resultado del IMERIN y en el marco de los propósitos de la Educación STEM.

Una vez cumplidos los objetivos se realizó una entrevista a cada una de las evaluadoras, con el fin de indagar y documentar sus conceptos, sugerencias e interpretaciones de los perfiles de habilidad inferencial y sus posibles usos en diversos ámbitos de la Educación Preescolar y la Psicología. La entrevista fue semiestructurada y permitió obtener información sobre supuestos relacionados con el origen de las diferencias en los conocimientos usados para guiar las inferencias. Además, se indagó sobre implicaciones de la variedad de desempeños y de la diversidad de las respuestas observadas. Puesto que, gran parte de los participantes hicieron conclusiones adicionales a las que abarcan los indicadores de eficacia, las cuales fueron plausibles y podrían ser elementos importantes en la caracterización del RI. Los conceptos se integraron a la descripción de perfiles y tendencias en la habilidad inferencial, en los resultados.

Participantes

La segunda evaluación fue realizada por 2 Psicólogas y una Licenciada en Educación Infantil, las 3 fueron convocadas por su amplia experiencia en el uso de herramientas para la evaluación diagnóstica, en áreas de la Psicología Clínica y de la Educación, o en la atención a la Primera Infancia en situación de vulnerabilidad socioeconómica. Para la conformación del grupo de evaluadoras se consideró, además, la conveniencia del concepto de profesionales representantes de usuarios potenciales de índices de desarrollo, de protocolos de evaluación y otros tipos de estrategias metodológicas para la valoración de capacidades cognitivas.

La participación de las evaluadoras consistió en repetir la calificación de los ítems de precisión, coherencia y relevancia respondidos por 26 (30%) de los 96 participantes de la aplicación para validación del IMERIN. Además de participar como entrevistadas, a las evaluadoras se les solicitó revisar las preguntas de la encuesta sobre prácticas de búsqueda de conocimiento en familia (en ese momento en versión de prueba) y dar su opinión sobre la utilidad de los resultados esperados de su aplicación.

Diseño

La segunda calificación de ítems y evaluación de desempeños siguió un diseño 3x2, de manera que cada una de las 3 evaluadoras evaluó, de forma independiente, 2 muestras de las aplicaciones en aula y a distancia (videos). Cada evaluadora recibió un conjunto diferente de aplicaciones: 10, 10 y 6 participantes respectivamente, estas fueron seleccionadas al azar y estaban conformadas por cantidades iguales de niñas y de niños (5 y 5, o 3 y 3), representantes de los 3 tipos de institución y de los 2 niveles en los que estaban inscritos la totalidad de los participantes.

Instrumento y recursos

Para la segunda evaluación se resumieron las guías de uso de los indicadores que fueron juzgadas por los expertos. Además, las evaluadoras usaron el manual de aplicación e interpretación de resultados del IMERIN, versión ajustada. Los contenidos de esta versión se referían únicamente a las subescalas de eficacia y presentaban una síntesis de las interpretaciones y propuesta de interpretación y uso de los puntajes. Las aplicaciones se presentaron en videos, éstos fueron editados de las grabaciones originales para que solo aparecieran registros de las respuestas y otros desempeños de los participantes, omitiendo las intervenciones de la primera evaluadora.

Las puntuaciones resultantes de la segunda evaluación fueron analizadas por las evaluadoras, conforme al manual de calificación seguido en la primera aplicación. Las medidas generadas por cada subescala de eficacia fueron usadas como criterio para esbozar un perfil de habilidades inferenciales de cada participante y un concepto sobre características destacadas y representativas del RI del grupo evaluado, cuando lo consideraron viable. Para los registros y documentación de las segundas evaluaciones se ajustaron los formatos usados en el juicio de expertos (Anexo H). Los cuestionarios usados para las entrevistas se crearon en formularios de Google Docs y la entrevista se realizó por videollamada, la documentación correspondiente se hizo mediante un resumen analítico.

Procedimiento

Inicialmente, se crearon criterios normativos para interpretar el significado de cada indicador de puntuación en las subescalas de eficacia y se consignaron en el manual de aplicación que se ajustó para este estudio. Estos criterios serían luego, el fundamento para el análisis del acuerdo interobservadoras (entre la primera y la segunda observadora). Las segundas calificaciones se analizaron mediante estadísticos descriptivos, para los análisis adicionales de cualidades psicométricas de las subescalas se obtuvo el porcentaje de acuerdos entre las puntuaciones y coeficientes de correlación Spearman.

Con el fin de probar hipótesis sobre la validez predictiva del IMERIN y la confiabilidad de la medida como criterio diagnóstico de habilidad, se estudiaron los atributos de la eficacia en el RI descritos por cada evaluadora. Se compararon las ubicaciones que asignó a los puntajes de los participantes en un nivel de habilidad inferencial determinado, usando las mediciones de eficacia que ella obtuvo e interpretó. Luego, a partir de estos conceptos y las demás sugerencias, se ajustó el manual de aplicación e interpretación de resultados del instrumento, incluyendo instrucciones detalladas para el administrador de la prueba y el acompañante del menor durante la aplicación no presencial.

Estudio 3: Diseño y aplicación de prueba del cuestionario sobre prácticas de socialización de la cognición en la infancia temprana

El objetivo de la indagación realizada mediante la encuesta fue validar hipótesis sobre expectativas, aspiraciones y prácticas parentales orientadas al aprendizaje conceptual de dominio específico y al desarrollo de habilidades de razonamiento de los niños, tales como la identificación

de nueva evidencia como fuente de conocimiento. Los siguientes son los principios que orientaron la creación de contenidos de la encuesta:

- La atención a ciertos dominios es moderada por preferencias individuales promovidas culturalmente.
- Diferencias entre grupos sociales se manifiestan en la diversidad de sus prácticas de búsqueda de conocimiento, aquel culturalmente aceptado.
- Diferentes inferencias, a partir de los mismos contenidos, son evidencia de la influencia del conocimiento deseable para el aprendizaje y de las prácticas de mediación del adulto en la formación del pensamiento del niño.

Etapa 3a. Tuvo por objetivo el diseño de contenidos del cuestionario sobre prácticas de socialización de la cognición, dirigida a las madres o acudientes de los niños que participaron en la aplicación para validación del IMERIN. Los contenidos de la versión inicial del cuestionario fueron evaluados por una profesora de Estadística experta en procesos de encuestas, en Colombia.

Contenidos de la encuesta. Una práctica de socialización de la cognición, estructurada, se caracteriza por integrar algún conocimiento de dominio específico, acciones pedagógicas concretas y expectativa de cambio conceptual, explícita o implícita. En síntesis, el cuestionario permite estimar la frecuencia de actividades representativas de 3 categorías de la mediación en la formación del pensamiento, estas son: a) comunicación de conocimiento y uso del lenguaje con intenciones de fomentar el desarrollo conceptual; b) búsqueda de experiencias para generar evidencia de causas y relaciones entre propiedades de interés; c) demostraciones pedagógicas de fenómenos en dominios específicos, para el aprendizaje de competencias “para la vida”. En la Tablas 8 y 9 se presenta la Matriz de especificaciones conceptuales para la creación de contenidos de la encuesta y la Tabla 10 presenta la descripción de las categorías de prácticas en las que se agruparon los ítems.

Tabla 8.

Matriz A de especificaciones conceptuales de los ítems de la encuesta sobre prácticas de socialización de la cognición.

Supuestos - Categoría búsqueda de conocimiento deseable	Antecedentes	Tipos de actividades familiares
<p>Las personas prefieren aprender sobre algún tipo de conocimiento específico</p> <p>Deseabilidad – valoración social- del conocimiento en dominios STEM</p> <p>La atención a ciertos dominios es moderada por preferencias individuales promovidas culturalmente</p> <p>Diferencias entre grupos sociales se manifiestan en la diversidad de las prácticas de búsqueda de conocimiento culturalmente aceptado</p>	<p>Nisbett et al 2001: La organización y las prácticas sociales afectan la plausibilidad de los supuestos de base para la inferencia de causalidad y otros procesos cognitivos</p> <p>Kushnir et al (2009) los niños consideran limitaciones en el conocimiento de un agente como criterio para evaluar causalidad</p> <p>Sloutsky et al, (2015) factores atencionales, como la atención a similitudes, influyen la inducción al inicio del desarrollo</p>	<p>Instrucción en solución de problemas</p> <p>Experiencias de descubrimiento de sistemas y relaciones de causalidad</p> <p>Actividades de indagación para la verificación de los hechos</p> <p>valoración de estrategias para el razonamiento eficaz</p>
Supuestos - Categoría comunicación del conocimiento	Antecedentes	Tipos de actividades familiares
<p>La comunicación del conocimiento es una práctica parental</p> <p>Es probable que exista valoración positiva de la mediación parental en la socialización de la cognición</p> <p>Atribución diferencial de relevancia a prácticas comunicativas, de verificación de conocimiento y revisión de creencias, propias y de los niños</p>	<p>Koenig et al, (2015) niños 3 años comprenden que el conocimiento verbalizado por el adulto es verificable</p> <p>Ruffman et al, (2012) los bebés son sensibles a las intenciones implícitas en la conversación de la madre</p> <p>Nisbett et al (2012)- Hablar con los niños fomenta la autorregulación del aprendizaje</p> <p>Clark, (2004); Mercer (2013) el ajuste entre los significados lingüísticos y las representaciones conceptuales es un factor de adquisición.</p>	<p>Aprendizaje de Representaciones semánticas y lingüísticas diversas</p> <p>inferencia de relaciones abstractas a partir de eventos cotidianos; descubrir, indagar y preguntar</p> <p>Sesgos o estereotipos asociados al género propio y de la hija -hijo</p>

Tabla 9.

Matriz B de especificaciones para la creación de ítems para la encuesta sobre prácticas de socialización de la cognición.

Supuestos - Categoría percepción de igualdad entre géneros	Antecedentes	Actividades familiares
<p>Existen diferencias en las interacciones sociales en las que las madres y padres buscan involucrar a los niños, en función de su género</p> <p>Diferencias de intencionalidad en la enseñanza de representaciones lingüísticas y modalidades de expresión de ideas en relación con el género del adulto y del niño</p>	<p>Kühhirt et al, (2018) el empleo materno mejora levemente oportunidades y recursos para el dlo cognitivo en la infancia temprana OECD 2015: madres y padres podrían tener creencias estereotipadas sobre las áreas en qué se destacan las mujeres y los hombres</p> <p>Lundberg 2020: creencias sobre la identidad de género influyen negativamente en las aspiraciones educativas de niños y niñas</p>	<p>Uso de interacción social en función del cambio en el conocimiento -conceptual; de categorías y relaciones de causalidad</p> <p>- facilitación de recursos TIC para la exploración de contextos STEM</p> <p>- diversidad en la interacción para la comprensión social del conocimiento en dominios específicos</p> <p>-Sesgos o estereotipos asociados al género</p>
Supuestos - Categoría experiencias de verificación del conocimiento	Antecedentes	Actividades familiares
<p>Es probable que los padres reconozcan capacidades de los niños de construir teorías intuitivas y revisarlas para verificar su relevancia</p> <p>Valoración de nueva evidencia como fuente de conocimiento</p> <p>Orientación en la atribución de sentido para experiencias y creencias, en función de la plausibilidad</p>	<p>Kushnir et al, (2009) experiencias de intervención en contextos ambiguos y con incertidumbre favorecen el aprendizaje causal en la edad preescolar</p> <p>Kuhn (2010) si las teorías de los niños son usadas para dar sentido a sus experiencias, ellos estarían construyendo ciencia implícitamente</p>	<p>motivación por la generación de conocimiento verificable</p> <p>Revisión de relevancia de las creencias propias y de los niños</p> <p>vínculo con ofertas culturales en temáticas científicas</p> <p>Sesgos o estereotipos asociados al género propio y de la hija -hijo</p>

Etapas 3b. El estudio de funcionalidad del cuestionario se realizó mediante una aplicación de prueba, o pretest que incluyó el autoreporte de prácticas. Además, se realizó una evaluación del formato y los encuestados aportaron opiniones de sobre la forma de ajustar los componentes, para facilitar el seguimiento de instrucciones y el correcto diligenciamiento del formulario. La prueba tuvo por objetivo general determinar la viabilidad del proceso de encuesta y específicamente,

verificar las suposiciones sobre el adecuado diseño, uso del lenguaje (considerado incluyente) y representatividad de las categorías de prácticas de socialización de la cognición que fueron consideradas. Se esperaba que la información suministrada por los evaluadores encuestados permitiera crear expectativas razonables sobre los resultados de la aplicación objetivo.

La aplicación de prueba del cuestionario se justificó, además, por la necesidad de conocer particularidades de la situación que favorecieran o dificultaran el acceso a dispositivos, redes en línea, o cualquiera de los recursos requeridos para autoadministrar el cuestionario. Dadas las restricciones al contacto físico durante el primer semestre de 2021, en el que se llevó a cabo el trabajo metodológico de la etapa 3. Los encuestados estuvieron a cargo de la gestión de la encuesta (autodiligenciamiento) y el envío de las respuestas se hizo a distancia mediante TIC, limitando así la intervención para apoyar la comprensión de requerimientos, objetivos e instrucciones de la participación. Por estas razones se justificó la aplicación de prueba.

Tabla 10.

Categorías de origen de los ítems de la Encuesta sobre prácticas de socialización de la cognición en la infancia temprana.

Item	Categoría: Prácticas de comunicación del conocimiento de la realidad
1 8 12 13 14 16	Actividades relacionadas con expresar creencias, preferencias, conceptos y significados compartidos por la comunidad y asociadas al uso intencional del lenguaje en la cotidianidad. Se supone que los adultos informan las denominaciones, nombres y otras claves lingüísticas que guían el RI de los menores y probablemente sean explícitas en las verbalizaciones de sus conclusiones (por ejemplo, Kim et al, 2012)
2 5 6 7 9 18	Categoría: Prácticas de búsqueda de conocimiento basado en la evidencia Acciones de búsqueda de experiencias para razonar y transformar conocimiento a partir de la observación, experimentación o descubrimiento de información, datos, covariaciones o interacciones entre elementos de diversas categorías. Se supone que los adultos intentan acercar a los menores al conocimiento en contextos de dominio específico, para que comprendan algunas leyes o principios relevantes de los fenómenos que éstos abarcan (por ejemplo, Buchanan & Sobel, 2011)
3 4 10 11 15 17	Categoría: experiencias de enseñanza-aprendizaje del conocimiento culturalmente valorado Actividades de formación y demostración de relaciones conceptuales que ponen a disposición del menor, recursos, medios y conocimientos que la cultura promueve con el fin de orientar el desarrollo cognitivo o intelectual. Se supone que las experiencias pedagógicas son creadas por el adulto con el fin de motivar al menor por el aprendizaje de competencias “para la vida”, en igualdad de oportunidades para niños y niñas, del siglo XXI (OECD, 2014; Butler & Tomasello, 2016).

Participantes

La aplicación de prueba del cuestionario sobre prácticas de socialización de la cognición fue realizada con la participación de 12 encuestados (y evaluadores), entre ellos: 3 profesoras con experiencia en enseñanza de las ciencias con población infantil de estratos socioeconómicos bajos; 8 madres de niños y niñas en edad preescolar, 3 de ellas profesoras en Primaria o Preescolar (además sus hijos participaron en la aplicación de prueba del IMERIN 1); y un padre de familia, profesor universitario y sociólogo. De los participantes, 11 representaban a los 6 estratos socioeconómicos de la población bogotana y 1 al sector rural de Cundinamarca. Sus niveles de estudios están entre secundaria y posgrado, 9 estaban empleados y 3 desempleadas. Ellos fueron contactados a través de las profesoras que participaron en la etapa 1 y las directoras de los Jardines que se interesaron en convocar la participación de sus estudiantes durante la etapa 3.

La participación fue voluntaria y consistió en: a) responder el cuestionario si tenían hijos en edad preescolar, o solicitar las respuestas a alguna persona que los tuviera (caso de 3 profesoras); b) evaluar la suficiencia o redundancia de los contenidos, de acuerdo con su experiencia en socialización de la cognición o en el acompañamiento al proceso académico en Preescolar durante el periodo de cierre de aulas; c) calificar la adecuada redacción y presentación de las opciones de respuesta, las instrucciones y el formato del instrumento; y d) escribir, o grabar sus opiniones sobre aspectos que debieran ser incluidos (o retirados) para facilitar la comprensión de los objetivos del proceso y el uso que tendría la información solicitada, o cualquier otro aspecto del cuestionario que consideraran relevante.

Diseño

Inicialmente, se convocó y vinculó a los evaluadores, posteriormente se les enviaron los materiales, se entrevistaron de forma individual y finalmente se interpretó la evaluación y conceptos que documentaron. La prueba se diseñó y realizó en las mismas condiciones planeadas para la aplicación objetivo, asignando los formularios según el género de los hijos del participante (o de los hijos de las personas que colaborarían con las profesoras) y enviando vínculos del formulario de Google Docs. por mensajes de texto a los correos y WhatsApp. El análisis de resultados del pretest fue descriptivo e interpretativo y se usó, principalmente, para modificar el

formato y extensión del cuestionario. Todo el proceso de participación y evaluación se registró por medios digitales y usando TIC.

Instrumento y materiales

Cuestionario de prueba. Para la versión preliminar del cuestionario se eligieron 3 categorías de hábitos, 21 ítems y 2 escalas numéricas con opciones entre 0 y 3 para elegir la respuesta, una de frecuencia y otra de importancia de la actividad referida en el ítem. Según el número de opción elegido, la respuesta indica lo siguiente:

- 0 significa que no realiza la actividad referida, o no considera de importancia el aspecto tratado en el ítem;
- 1 significa baja frecuencia de la actividad durante las últimas 4 semanas (realizada pocas veces), o poca importancia de la actividad;
- 2 corresponde a una frecuencia media de la actividad (varias veces se realizó), o a un nivel medio de importancia; y,
- 3 se refiere a una frecuencia alta de la actividad (un hábito), o a un alto nivel de importancia del referente.

Los 21 ítems creados, las instrucciones para el diligenciamiento y la descripción de la importancia y uso previsto de la información derivada de las respuestas quedaron escritas en un formato único, usando la herramienta Google Forms. Para consignar los resultados de la evaluación se crearon materiales como listas de verificación, formatos con indicaciones y espacios para el registro de las observaciones. Los participantes realizaron registros de audio que se resumieron por medio de tablas de especificaciones (en físico).

Procedimiento

Una vez se adecuaron los ítems, según el juicio de la experta en encuestas, se realizó la convocatoria y vinculo de los evaluadores participantes en la prueba de viabilidad y funcionamiento del instrumento. Posteriormente se realizó una entrevista individual a cada evaluador, esta fue semiestructurada (por videollamada) y tenía por objetivos confirmar y retroalimentar la comprensión de las indicaciones para la participación.

La aplicación de prueba fue autoadministrada, siguiendo un plan concreto de dos fases, distribución y recopilación de las encuestas, así como un cronograma de corta duración. La interpretación de los conceptos de los evaluadores -encuestados se enfocó en verificar el

cumplimiento de objetivos. Algunos de los cambios sugeridos se incorporaron a la versión final de la encuesta, estos se detallarán en el capítulo de resultados.

Evaluación de viabilidad de la autoaplicación de la encuesta mediante TIC. Los 12 evaluadores coincidieron en que por medio de la encuesta se lograría obtener la información pertinente para comprender algunos hábitos y creencias subyacentes a las acciones concretas de socialización del conocimiento. Así mismo, que los formatos y procesos para conseguir dicha información eran viables y fáciles de comprender. Incluso, estuvieron de acuerdo en que la autoadministración del cuestionario favoreció la espontaneidad y comodidad del encuestado para responder, afirmando sentirse en mejor disposición para la participación en comparación con situaciones en las que la administración fue controlada por un entrevistador, o si se requería interactuar con el administrador del cuestionario. En conclusión, a su juicio, las expectativas sobre la funcionalidad del instrumento en relación con lograr los propósitos de la aplicación objetivo fueron consideradas razonables.

De otra parte, las encuestadas que habían presenciado la aplicación de prueba del IMERIN-P, una vez que respondieron el cuestionario, reconocieron diferentes influencias en el RI de sus hijos de las prácticas de formación del pensamiento realizadas, incluso si éstas fueran ocasionales. En concordancia con la idea de Ebersbach & Resing (2008), sobre el conocimiento implícito de los menores como el origen de inferencias espontáneas, las encuestadas afirmaron que practicar la generación de conclusiones de sus hijos (la inferencia), les permitiría entender el tipo de conocimiento conceptual del que disponen para pensar “autónomamente” (sin intervención de un adulto) y cómo influir en su desarrollo directamente.

En ese orden de ideas, los encuestados manifestaron haber interpretado que el logro en la formación de hábitos de pensamiento tiene al menos una forma concreta de controlarse; por medio de cambios en la frecuencia de las prácticas de socialización de la cognición. Esta coincidencia entre la expectativa de los evaluadores sobre la posibilidad de valorar, con criterios objetivos, sus metas en el acompañamiento y el supuesto teórico del estudio, indica coherencia entre el fundamento teórico de los contenidos de la encuesta y la interpretación de resultados proyectada. El supuesto se refiere a la pertinencia de usar la frecuencia de prácticas de socialización como magnitud de la influencia social en el desarrollo cognitivo y se basa en las ideas de Ruffman et al,

(2012). Es importante mencionar que los fundamentos conceptuales no se especificaron en ninguno de los documentos de orientación para la participación.

A partir de la evidencia referida se consideró razonable esperar que los participantes de la aplicación objetivo encontraran en la encuesta herramientas para controlar, hacer seguimiento y evaluar un aspecto favorable del proceso de involucramiento que han realizado durante la desescolarización. Por ejemplo, el 80 % de las madres encuestadas manifestaron intenciones de incrementar la frecuencia de ciertas prácticas referidas en los ítems para que fueran hábitos, e interés en indagar con las profesoras de sus hijos sobre la creación de actividades que mejoren la mediación de las familias en el aprendizaje de los menores, de contenidos en dominios STEM y especialmente, de las prácticas relacionadas con controlar el riesgo de infecciones y el cuidado del sistema inmunológico, incluso después del retorno a la escolaridad.

Ajustes al instrumento y procedimiento de encuesta. Los evaluadores confirmaron que 18 de los 21 ítems tienen un diseño adecuado, que facilitan comprender las instrucciones para el diligenciamiento y la importancia de la información aportada en las respuestas. Respecto a redacción y forma, consideraron que el uso de lenguaje amable e incluyente favorece la vinculación del encuestado al logro de objetivos del estudio. Respecto a su propia experiencia, les pareció adecuado informar objetivos y dar instrucciones a los participantes durante una corta entrevista, posterior a la videollamada de aplicación del IMERIN, confirmando y retroalimentando la comprensión de las indicaciones para la participación.

Etapa 3c: Aplicación de la encuesta

La justificación de la etapa 3c radica en la necesidad de generar evidencia empírica (ecológicamente válida) para usar como referente en la exploración hipótesis sobre expectativas, intenciones y prácticas parentales orientadas al aprendizaje de conocimiento de dominio específico y al desarrollo de habilidades de pensamiento de los menores participantes. Por medio del proceso de encuesta se estudiaron dos variables fundamentales de la socialización: *frecuencia* de las actividades cognitivas compartidas, e *interés* en la adquisición del conocimiento percibido como socialmente deseable para la formación del pensamiento en la infancia. Además, el instrumento indaga si el interés por desarrollar conocimientos de dominios STEM difiere en relación con el género del menor participante.

Participantes

En este estudio participaron 80 madres y 1 padre, es decir, 81 de los 96 acudientes de los menores que participaron en la etapa 3b, entre ellos 30 son madres de niñas y 51 de niños. De los 81 participantes, 48 estaban empleados o recibían un ingreso económico mensual y 33 estaban desempleadas, 19 tienen título profesional, 20 tienen estudios técnicos o tecnológicos, 41 son Bachilleres y 1 estudió Primaria. Además de responder la encuesta, la participación consistió en acompañar al menor durante la aplicación del IMERIN-P, facilitando el acceso a los medios necesarios para realizarla, sin intervenir. Posteriormente las madres fueron entrevistadas (por segunda vez) para atender sus inquietudes o sugerencias y brindarle información sobre el avance en el proceso de investigación y sus repercusiones en las actividades planeadas por las profesoras de los niños.

A pesar de que los 96 acudientes de los participantes del estudio 2 habían aceptado participar con sus hijos, 15 de ellos no respondieron el cuestionario. La mayoría de ellos informaron que las razones de no participar se referían a consecuencias directas de la Pandemia. Las personas vinculadas al jardín privado que no acompañaron la aplicación al IMERIN-P, fueron un porcentaje amplio de quienes no participaron. Algunos explicaron que habían olvidado el cuestionario, dadas sus ocupaciones de trabajo (aunque se les recordó en varias ocasiones). Otra razón frecuentemente aludida fue la falta de acceso al servicio de Internet. El abandono en este estudio podría asociarse, incluso, con sentirse confrontados por las preguntas, como lo creyeron un par de directoras de las instituciones participantes.

Diseño

El estudio 3c fue exploratorio, por sus objetivos y transversal de acuerdo con la dimensión temporal del proceso indagado, de los datos buscados y de sus fuentes de información (estimación de frecuencias durante un periodo determinando). La frecuencia de la actividad de socialización cognitiva (o práctica) fue usada como indicador de hábito y los hábitos que se asociaron con puntajes de los niños en las dimensiones de eficacia del IMERIN. Luego fueron incluidos en el análisis de resultados, como características relevantes del desarrollo del RI en algunos grupos de participantes. Para la aplicación del instrumento se realizó un muestreo por conveniencia, no probabilístico.

La técnica de ordenamiento y sistematización de las respuestas cerradas es un criterio incluido en el diseño del formulario de Google Docs., así que esta parte del procesamiento se automatizó, facilitando el resumen de los datos obtenidos. Se usaron herramientas de la Estadística descriptiva y análisis de correspondencias simples para analizar agrupaciones de los datos, se obtuvieron coeficientes de correlación Sperman entre las prácticas de alta frecuencia y los puntajes totales del IMERIN y se realizó la prueba estadística Kruskal-Wallis para probar hipótesis sobre la asociación entre diferencias en los niveles de eficacia en la inferencia de los menores y algunos hábitos indagados. Todos los procesos estadísticos se realizaron mediante herramientas del entorno R.

Instrumento

Encuesta sobre prácticas de socialización de la cognición en la infancia temprana. El principal instrumento de la encuesta es el formulario digital que contiene los 18 ítems del cuestionario inicial, las instrucciones para el diligenciamiento y la descripción del uso previsto para información aportada por los encuestados. El cuestionario está conformado por: 15 ítems que indagan frecuencias de las prácticas cognitivas y tienen múltiples opciones de respuesta; 2 ítems en los que se pregunta por el área de conocimiento de interés personal y la que esperan que se formen los niños; y, 1 ítem que pregunta por la percepción de igualdad en el acceso al conocimiento en la infancia, en función del género (se solita explicación). EL cuestionario ésta en el Anexo Q.

Procedimiento

El proceso que permitió encuestar a las madres inició con la convocatoria facilitada por las docentes que participaron en el estudio 1. Un alto porcentaje de los convocados se interesaron en participar y diligenciaron el consentimiento informado, en éste se explica, en términos generales, los objetivos y forma de participación de los encuestados. Posteriormente se realizó una entrevista individual por medio de llamada, en la que se confirmó su interés y se resolvieron inquietudes sobre el proceso. Una vez se determinó la agenda para vincular al grupo de niños participantes en el estudio 2, sus acudientes fueron integrados como participantes del estudio 3.

Se realizaron tres comunicaciones, de corta duración, con las participantes; en la primera se les recordó los objetivos de la participación (incluyendo la de los menores), el tratamiento confidencial de la información reportada. Así como su utilidad para la investigación, los intereses pedagógicos (y diagnósticos) de las docentes y de la institución que convocó. También se confirmó

su interés en participar, aunque previamente habían diligenciado el documento de consentimiento informado. La segunda entrevista se realizó al finalizar la aplicación del IMERIN-P con los menores, en esta se verificó y retroalimentó la comprensión de las indicaciones para habilitar, diligenciar y devolver el formulario resuelto, inmediatamente se envió el vínculo a los dispositivos y correos. Recibidas las respuestas, se les entrevistó por última vez, con el fin confirmar el recibido, agradecer su participación, atender inquietudes, ofrecer sugerencias sobre actividades que podrían realizar con los menores para fomentar la eficacia en el RI y brindar información, si se solicitaba.

Capítulo 3: Resultados

Debido al cambio en parte de la metodología de investigación, los resultados del estudio 1 se integraron a resultados de algunas etapas del estudio 2. Estos se resumieron en la descripción de métodos, en el capítulo 2.

Estudios de validación del IMERIN y su utilidad científica

Estudio 2a. Validación de contenidos por Juicio de Expertos

Con el objetivo de obtener evidencia de la validez de los contenidos del IMERIN-P se realizó un juicio de expertos. Ellos juzgaron la coherencia teórica y metodológica de la instrumentalización como estrategia para obtener datos confiables sobre las características del RI en la infancia temprana. Además, evaluaron la pertinencia, suficiencia, objetividad y redacción de los ítems de medición de eficacia y sus indicadores de precisión, coherencia y relevancia. El resultado del juicio implica que el IMERIN es un instrumento válido y confiable para la medición de la eficacia del RI de los niños con edades entre 4 y 6 años. Por esta razón, la interpretación de los resultados del instrumento aporta confiabilidad a la propuesta metodológica para la caracterización objetivo.

Plan de análisis. El resultado del juicio se analizó a partir de información estadística sobre acuerdos entre los jueces en el rango de calificación para cada una de las 9 actividades, además se interpretaron sus conceptos individuales. El indicador de acuerdo interjueces fue analizado a partir de la comparación del porcentaje de la calificación máxima posible que se asignó cada juez en cada actividad. Se obtuvo también este dato por cada ítem, adicionando y promediando las calificaciones de los 4 jueces en las 4 categorías de juicio. Es importante mencionar que la estrategia para obtener

el índice de acuerdo se justificó dado el tamaño de muestra de los ítems y el número de jueces (ver aclaraciones sobre el coeficiente ω de Kendall en Bonett y Wright, 2000). (Anexo I.)

Calificaciones de los ítems de medición de la eficacia en el RI.

Se analizaron los resultados cuantitativos del juicio a partir de la adición de las calificaciones dadas a los 27 ítems de las dimensiones de eficacia juzgadas, usando la estrategia descrita por Corral, (2009) denominada “agregados individuales”. En la Tabla 11 se observan los puntajes asignados por los 4 jueces al conjunto de 3 ítems, en las 4 categorías de validez establecidas (pertinencia, suficiencia, objetividad y redacción). El rango de calificación del ítem estuvo entre 0 y 3, siendo 48 puntos la máxima calificación posible para el grupo en cada modalidad de inferencia. En cada actividad en la que no hubo acuerdo total se analizó la observación registrada por el juez en desacuerdo, ajustando en cada caso, según la recomendación.

También se obtuvo el porcentaje de acuerdos interjueces en los puntajes de cada categoría de juicio. Ambos resultados, en especial los promedios de objetividad, así como los conceptos y sugerencias de los jueces, se usaron para ajustar los indicadores de las 6 actividades elegidas para conformar la versión final del IMERIN (el instrumento que se aplicó para validación, Figuras 12 a 30).

Respecto a la pertinencia y adecuada redacción de los indicadores de precisión de las dos actividades del dominio Ciencias, el acuerdo interjueces fue del 100%; así mismo, para los indicadores de suficiencia de la actividad de explicación de este dominio. En cuanto al dominio Estadística, en la actividad de explicación hubo 100% de acuerdo en la suficiencia y adecuada redacción de los indicadores, así mismo, para la actividad de predicción en la categoría pertinencia. En Tecnología, se dio 100% de acuerdo en la adecuada redacción de los indicadores de la actividad de explicación.

Tabla 11.

Puntajes totales de los jueces para las dimensiones de eficacia - IMERIN-P (N=27).

Dimensiones de la eficacia en el RI					
	Modalidad	Actividad	Precisión	Coherencia	Relevancia
Ciencias	Hipótesis	1	45	44	41
	Explicación	2	47	45	48
	Predicción	3	47	48	47
Estadística	Hipótesis	4	25	25	27
	Explicación	5	46	46	47
	Predicción	6	47	44	46
Tecnología	Hipótesis	7	38	40	43
	Explicación	8	48	47	45
	Predicción	9	46	41	43

Los porcentajes de acuerdo en objetividad de los indicadores en las actividades de los tres dominios fueron superiores al 93%. No obstante, se realizaron los ajustes sugeridos por los jueces que calificaron con menos de 3 en esta categoría. Las sugerencias estaban enfocadas a incluir indicadores adicionales para comparar la conclusión, o en mejorar la discriminación entre éstos. También se cambiaron imágenes y se mejoró la redacción de los indicadores en la actividad de predicción en Tecnología, siguiendo las indicaciones de los jueces, aunque el porcentaje de acuerdo para la calificación de los indicadores en las 4 categorías de juicio fue superior al 80%. La Tabla 12 presenta las calificaciones promedio en los indicadores de las 6 actividades elegidas para conformar el IMERIN validado.

A partir de las calificaciones totales, se puede afirmar que los ítems que mejor miden la precisión son los de la actividad explicación, en el dominio Tecnología (obtuvo la calificación máxima de 48 puntos) (Figura 28). Así mismo, calificaron con el máximo puntaje los ítems de coherencia en la actividad de predicción y en la dimensión relevancia en la actividad de explicación del dominio Ciencias (Figuras 12 a 17). También en el nivel de mejores calificaciones están los ítems de la actividad de explicación y predicción en Estadística (Figuras 18 a 26) y los ítems de coherencia y relevancia de la actividad de explicación en Tecnología (Figuras 29 y 30).

Tabla 12.
Calificación Promedio en los indicadores de las 6 actividades del IMERIN (N=36)

Actividad	Categorías	Calificación Promedio en los indicadores			
		Pertinencia	Suficiencia	Objetividad	Redacción
	Juez 1	3	3	2.33	3
Explicación	Juez 2	3	3	3	3
Ciencias	Juez 3	3	3	3	3
	Juez 4	3	3	3	3
	Juez 1	3	3	2.66	3
Predicción	Juez 2	3	2.66	3	3
Ciencias	Juez 3	3	3	3	3
	Juez 4	3	3	3	3
	Juez 1	3	3	2	3
Explicación	Juez 2	2,66	3	2.66	3
Estadística	Juez 3	3	3	3	3
	Juez 4	3	3	3	3
	Juez 1	3	3	2.33	2
Predicción	Juez 2	3	3	3	3
Estadística	Juez 3	3	3	3	3
	Juez 4	3	2.66	3	3
	Juez 1	2,66	2,66	2,33	3
Explicación	Juez 2	3	3	3	3
Tecnología	Juez 3	3	3	3	3
	Juez 4	3	3	3	3
	Juez 1	2,33	2,66	2,66	2
Predicción	Juez 2	3	2	3	2,33
Tecnología	Juez 3	2,33	3	3	3
	Juez 4	3	3	3	3

La actividad de hipótesis en Ciencias obtuvo buenas, así como los ítems de la actividad de hipótesis en el dominio Tecnología. Sin embargo, las dos actividades se retiraron del instrumento. Dado que, en el juicio se estableció que la evaluación de hipótesis estaba implicada en las modalidades explicación y predicción, por lo cual las mediciones adicionales resultarían redundantes. Los 2 jueces que plantearon este concepto recomendaron, además, reducir las actividades con el fin de aplicar el instrumento completo en una sola videollamada. Por esta razón,

la actividad de hipótesis en Estadística también se retiró. Además, obtuvo calificaciones bajas en las 3 dimensiones, esto indica que los jueces no estaban de acuerdo en validar su pertinencia, suficiencia y objetividad para la medición requerida. Posteriormente, el retiro de las tres actividades de hipótesis facilitaría la estandarización de los puntajes.

Estudio 2b. Características psicométricas y resultados de la aplicación para validación del IMERIN

La aplicación del instrumento se justificó por la necesidad de validar la propuesta de interpretación y utilidad científica de los resultados del IMERIN. Por esta razón, el objetivo del estudio fue generar evidencia empírica, adicional, sobre el uso del instrumento validado para la evaluación de diferencias individuales y características similares en los atributos de las inferencias eficaces a nivel de grupos. Es decir, la prioridad fue probar supuestos sobre el funcionamiento de los ítems, generando evidencia estadística de la confiabilidad y de validez de la interpretación de puntajes totales y demás resultados que generan los componentes del IMERIN. Además, se requería obtener estadísticos de confiabilidad del procedimiento de aplicación. Dado que el estudio de propiedades psicométricas tiene el propósito de sustentar que la instrumentalización, con fines de medición de eficacia en el RI, es fundamental en su caracterización. En este caso particular se debía probar, además, equivalencia de la medición en las dos modalidades de aplicación (en aula y a distancia).

Plan de análisis. Esta parte del estudio hizo énfasis en verificar que los participantes en efecto respondieran los ítems según lo esperado, independientemente de las particularidades de los medios, o modalidades de aplicación del instrumento. Se esperaban altos índices de coincidencia entre las conclusiones de los participantes y los indicadores, en ambas modalidades de la evaluación (en aula -presencial- y a distancia -no presencial-).

La aplicación se realizó con la participación de 108 niños, los puntajes de 12 de ellos no se incluyeron en el análisis de resultados porque abandonaron la aplicación antes de terminarla, o no obtuvieron calificación 1 en tres ítems seguidos, por haber superado los 6 años y 1 mes y por pérdida del contacto con el acudiente que firmó el consentimiento informado. De manera que el análisis se hizo con una muestra de 96 conjuntos de datos ($N= 96$) provenientes de: 46 participantes con edades entre 4 años y 1 mes, hasta 4 años y 11 meses; 43 participantes de 5 años y 0 meses, hasta 5 años y 11 meses; y, 7 participantes con edades superiores a 5 años y 11 meses, hasta 6 años

y 0 meses (la descripción del tipo de Institución vinculante se realizó en el capítulo anterior). Para realizar estos análisis, se utilizaron diferentes paquetes del entorno R (versión 4.1.0, 2021) y Excel (versión 2016).

Para verificar los supuestos del estudio, se analizó exhaustivamente el funcionamiento de cada ítem, en términos del desempeño de los participantes en cada uno de los dominios. El desempeño adecuado se trató como acierto en la conclusión, éste depende de la coincidencia con indicadores. En este orden de ideas, se requería hacer una descripción estadística de aciertos para obtener evidencia inicial de que los participantes hicieron juicios generalizantes plausibles, inferencias precisas, coherentes y relevantes en las modalidades explícitas observadas. Se debía analizar si estas respuestas eran acordes con las características del contenido de cada indicador y con las particularidades de la evidencia en cada ítem. Es decir, se estudió el comportamiento de las respuestas o conclusiones en relación con los supuestos sobre funcionalidad de la estructura inferencial modelada por las situaciones de RI.

Actividad Tecnología –predicción

Objetivo: evaluar el RI mediante la predicción de resultados conocidos en nuevas condiciones, parcialmente desconocidas. Integrando información sobre control de variables representativas de diferentes dominios. Implica inferir la repetición de patrones de relación conocidos a pesar de los cambios en las condiciones del proceso, por ejemplo: uso de nuevos instrumentos y nuevas tecnologías.

Contexto: la agencia humana podría ser imitada por tecnologías, así como su capacidad de aprender técnicas de preparación de alimentos. La probabilidad de llegar al mismo resultado, que en este caso es lograr la transformación de ingredientes en comidas, depende de seguir un plan establecido.

Fase inicial familiarización, objetivo: generalizar la relación entre aprender secuencias para cocinar, usar recursos y métodos, y lograr el objetivo planeado; es decir, preparar comidas



Las niñas están aprendiendo a usar los instrumentos de la cocina, ¿Cuáles serán los resultados que conseguirán?

Indicadores de inferencia de la relación generalizada

- Las niñas observaran cambios de estados de los alimentos por la acción de cortarlos, cocinarlos, mezclarlos o cualquiera de las posibles intervenciones sobre estos
- Las niñas cumplirán su deseo de comer alguna comida favorita que les están enseñando a cocinar
- Las niñas lograran preparar un almuerzo o cualquier tipo de comida que suplirá su necesidad de alimentarse

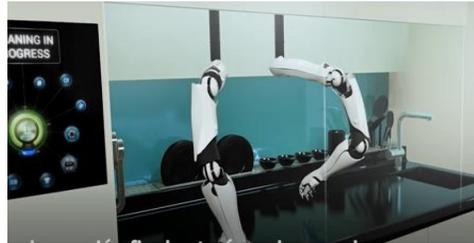
El criterio para pasar a la siguiente fase es que se verifique 1 de 3 indicadores de la generalización plausible

	Respuestas de algunos participantes	Cal.
1	Aprenden a hacer ensaladas	1
2	Harán un desayuno, pero toca tener cuidado con el cuchillo	1
3	Las personas adultas están por ahí para decirles y luego hacen su almuerzo	1
4	Van a cocinar	1
5	Quieren cocinar una comida	1
6	Para hacer una comida	1

Justificación de la calificación del ítem a partir del uso de los indicadores de plausibilidad
1 Generaliza el tipo de cambio más común en la clase de alimentos que se están manipulando, las ensaladas.
2 Generaliza el resultado de las acciones observadas, especificando una condición que representa diversidad en las categorías, los niños usualmente no usan cuchillos.
3 Generaliza la relación enseñanza – aprendizaje implícito en la introducción del caso y el resultado de la interacción inferida. Usa conocimiento de la condición excepcional de la escena; son niños cocinando, por lo cual infiere que debe haber presencia de los adultos.
4 Generaliza el caso distintivo de la relación usar utensilios de cocina para lograr un resultado
5 Generaliza un evento social asociado a las conductas evidentes en la imagen
6 Generaliza el resultado de usar utensilios de cocina, alimentos y acciones típicas en la cocina

Objetivo fase de prueba: inferencia de nuevo caso de aprendizaje de métodos para conseguir resultados, por integración de la tecnología, la voluntad y el conocimiento de técnicas de cocina.

Ítem de precisión: Este robot fue creado para ayudar a las personas ...



¿Cómo se podría enseñar a cocinar a un robot?

Indicadores de Precisión

Prever que deberá darse instrucciones, seguir recetas, o mostrarle cómo usar utensilios, electrodomésticos e ingredientes

Anticipar que los robots aprenden a transformar alimentos en comidas si "se les enseña" un paso a paso, o una técnica, o una secuencia

Respuestas de algunos participantes	Cal
1 Le pones con los botones instrucciones de cocina, de internet se miran recetas	1
2 Fue creado para ayudar a cocinar y lavar, con una computadora, aquí está	1
3 Que si lo escribes en su pantalla, como se echan los ingredientes, ¿si sabe prender la estufa?, tocaría que mire	1
4 Pero no le veo la cabeza	0
5 Le espichas así los botones	0
6 Porque es un robot de jugar	0

Fase de prueba Ítem de coherencia

Introducción: este es otro robot que ayuda a las personas en su trabajo..



¿Para qué está usando la cocina este robot?

Indicadores de Coherencia

Anticipar que las acciones del robot tienen el objetivo de cocinar alimentos para las personas

Proponer que los alimentos que prepara el robot serán la comida que solicitó el dueño del robot y/o de la cocina

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia

- 1 Infiere el principio causal correcto, programar instrucciones para seguir un proceso, usando conocimiento conceptual y la evidencia directa e indirecta
- 2 Infiere el principio de la agencia humana implicado en la función razonada, usando la evidencia directa y su conocimiento sobre ésta
- 3 Infiere el proceso que debe seguirse para conseguir el resultado, anticipa acciones necesarias haciendo uso de su conocimiento sobre las categorías y relaciones entre elementos del sistema que conducirán al resultado
- 4 Presentó dificultades para razonar haciendo uso de la evidencia, dirigiendo su atención a las características de la apariencia que creyó están asociadas con la tecnología presentada
- 5 Usa la evidencia y su conocimiento de las categorías, pero éste es impreciso e incompleto
- 6 Expresa una creencia sesgada respecto a las relaciones que asocia con el uso de elementos de la categoría tecnológica que conoce; el uso lúdico de los robots

	Respuestas de algunos participantes	Cal.
1	Para ayudarlo a las personas a hacer sus comidas	1
2	Está haciendo un almuerzo o una comida	1
3	Lava loza, como yo ...cocina de todo	1
4	Para aprender a cocinar Yo quiero enseñarle ...enseñándole	1
5	No sé	0
6	Esos son los brazos del robot.... no sé qué hace	0

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia
1 Infiere el objetivo de la relación a partir de la evidencia y la conclusión anterior
2 Infiere a partir de su conocimiento de las categorías y relaciones presentadas
3 Infiere a partir de su conocimiento de las labores cotidianas asociadas a la evidencia
4 Infiere a partir de las conclusiones anteriores y su conocimiento de la agencia humana
5 y 6 No usan evidencia ni conclusiones antecedentes

Ítem de relevancia



Introducción: Esta es otra máquina para preparar alimentos...

Ítem 3 – Relevancia: ¿Cómo podría aprender la niña a usar esta máquina?

Indicadores de Relevancia

Proponer que repetirá algunas actividades de enseñanza y aprendizaje, o técnicas para cocinar que ya conoce, asignando funciones o intervenciones al robot

Anticipar que usaría el robot de manera similar a como se han usado otras máquinas y electrodomésticos de la cocina

	Respuestas de algunos participantes	Cal.
1	Si mira las recetas que tiene el aparato y con la mamá le enseña cocinando	1
2	Ahí lee, que toca echarle y la mamá le ayuda conectando, diciéndole como	1
3	El papa le enseñó, mirándole ahí en el tablerito y echándole la leche....	1
4	Toca que mire la máquina	0
5	Porque ella cocina	0
6	Si la mamá el enseña, con las recetas y ponen todo adentro y la prenden	1

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia
1 Infiere el resultado integrando su conocimiento de las categorías y nueva información sobre la relación que predice el resultado
2 Infiere el proceso necesario para llegar al resultado agregando información y usando evidencia implícita
3 Infiere interacciones que favorecen llegar al resultado agregando información y usando evidencia implícita
4 y 5 No infieren el proceso necesario para llegar al resultado indagado
6 Infiere el proceso requerido para conseguir el resultado, agregando información presentada e inferida en los ítems pasados

Actividad estadística

Objetivo: evaluar la habilidad de predicción diagnóstica, a partir del RI sobre múltiples riesgos de contagio con el virus Covid-19. Implica usar conocimientos sobre categorías y sistemas de causalidad recientemente aprendido. Si se infiere la interacción de causas suficientes vs. inhibitoras es posible predecir la probabilidad de que los cuidados prevengan el contagio. Además, podría inferirse que, es probable que las conductas de riesgo se realicen en conjunto con las de cuidado, en este caso, también se podría predecir que habrá contagio.

Contexto: Durante el año 2020 y 2021 aparecieron nuevas fuentes de evidencia y se incrementaron los medios para publicar y poner a disposición, información y datos estadísticos sobre el curso de la Pandemia por Covid-19. De manera que es probable que las personas aprendieran nuevo conocimiento conceptual (de categorías y causalidad). Este sería relevante para el RI sobre el riesgo de desarrollar enfermedades, probabilidades de contagio, de evitación de contagio, manejo de la enfermedad y conductas de cuidado de la salud.

Fase inicial Familiarización



Introducción: Ahora debemos usar mascarilla y lavarnos las manos varias veces al día, ¿sabes por qué está ocurriendo esto?

Objetivo fase inicial: generalizar el riesgo de contagio, usando conocimiento de causas generadoras e inhibitoras, recientemente aprendidas. Revisando las creencias sobre cuidado de la salud en relación con nueva evidencia, indirecta, de riesgos y consecuencias de la exposición al virus

¿Si las personas se alimentan saludablemente y hacen ejercicio, ¿se pueden contagiar con coronavirus?



Indicadores de generalización plausible

- Suponer que las conductas de cuidado de la salud no son iguales a las conductas de prevención identificadas en las imágenes de la fase de familiarización
- Proponer que los contagios son posibles en todas las personas que no sigan las recomendaciones de prevención, aunque se alimenten bien y hagan ejercicio
- Afirmar que todas las personas estamos en riesgo de contagio

	Respuestas de algunos participantes	Cal.
1	Todos los que salgan a la calle sin tapabocas nos podemos enfermar	1
2	No sabemos si todos ya están contagiados, pero si usamos tapabocas de pronto no nos contagiemos....	1
3	Si comemos frutas muy bien, el brócoli ...si nos vamos a la calle sin tapabocas nos podemos enfermar por los virus	1
4	Si, los virus están por todos lados, peor si no comemos las vitaminas	1
5	Sí, porque no se ponen tapabocas o se lo quitan para comer	1
6	Si, los que coman frutas, verduras... si no se lavan, se tocan la cara y así se les prende el covid, en la calle, si van en el bus, o donde sea	1

Justificación de la calificación del ítem a partir del uso de los indicadores de plausibilidad

- 1 Generaliza el resultado de exponerse a las causas del contagio usando evidencia conocida
- 2 Generaliza su hipótesis frente a las condiciones que causan el contagio, agregando conocimiento de causas inhibitoras
- 3 Generaliza usando la evidencia para afirmar su conocimiento de las relaciones presentadas
- 4 Generaliza las condiciones causales usando su conocimiento de las categorías
- 5 Generaliza el efecto de exponerse al riesgo durante las actividades en la situación
- 6 Generaliza las causas del contagio expresando su conocimiento de la relación indagada

Objetivo fase de prueba: inferir nuevos casos de contagio en diversas condiciones de exposición y con diferente probabilidad de funcionamiento de las medidas de prevención conocidas



Ítem de precisión

Introducción: Esta familia fue a pedir que les hicieran exámenes para saber si están contagiados, porque creen que algunos de los niños se tocan la cara...

- ¿Aunque usen mascarilla y guantes, les harán los exámenes para saber si están contagiados?

Indicadores de Precisión

Anticipar que a todos les harán exámenes porque la conducta que se describió si implica riesgo de contagio

Suponer que no se sabe quiénes descuidaron las demás acciones de prevención, a pesar de que están usando mascarilla y guantes, por lo que deben hacer la prueba a todos

	Respuestas de algunos participantes	Cal.
1	Si les hacen, a todos si no, después que unos se les prendió de pronto a todos	1
2	No les hacen examen porque no salieron a la calle	1
3	Toca que les hagan exámenes de covid	1
4	Si se los hacen, porque salieron sin tapabocas	1
5	Pues unos si se tocaron, otros no se saben, pero si el bebé que no tiene tapabocas, yo creo que si se les hicieron a todos y la mamá también	1
6	Si les toca hacerse a todos, porque están en el hospital	1

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia

- 1 Infiere la condición necesaria para el conocer el estado de la situación, descartando información irrelevante
- 2 Infiere ausencia de causas para el evento indagado, usando su conocimiento de las condiciones que aumentan la probabilidad de ocurrencia
- 3 Infiere el resultado exacto de la situación indagada
- 4 Infiere las condiciones necesarias que justifican alta probabilidad de ocurrencia del evento razonado
- 5 Infiere circunstancias que aumentan la probabilidad de ocurrencia, adicionando creencias relevantes para el caso
- 6 Infiere la probable influencia de causas suficientes, que no son evidentes, usando su conocimiento de los contextos en los que se presenta la situación

Fase de prueba Ítem de coherencia



Introducción: Antes del examen la Enfermera le preguntó a la mamá: si ellos han compartido juguetes, si el baño está sucio, si salieron a la calle y se acercaron a otros niños que estaban tosiendo o estornudando

¿Por qué la Enfermera le pide esas respuestas a la mamá?

Indicadores de Coherencia

Predecir que la información se usará para decidir si deben examinar a todos o solo algunos, según quienes realizaron conductas de riesgo y quienes no las realizaron

Anticipar que la enfermera conseguirá información importante para sospechar que todos están contagiados y los enviaran a hacerse exámenes

Respuestas de algunos participantes		Cal.
1	Porque son muchos niños y todos salieron, o no se sabe si estaban enfermos	1
2	Porque ya sabemos que toca hacerles exámenes y la mamá los acompañó	1
3	Para saber si unos no los van a examinar y otros si toca que les sacan sangre	1
4	Porque salieron sin tapabocas	0
5	Porque la mamá sabe cuáles se quitan el tapabocas, o si están con virus	1
6	Para sacarles los exámenes	1

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia
1 Infiere la necesidad de tener información adicional dadas las condiciones de probabilidad del evento
2 Infiere las razones suficientes para el suceso, usando las conclusiones anteriores y la nueva evidencia
3 Infiere el suceso dadas las probabilidades, usa información de la situación y sus teorías
4 Repite respuestas anteriores desatendiendo a la evidencia
5 Infiere la importancia del conocimiento de causas probables para el evento y sus fuentes, usando su conocimiento de evidencia presentada anteriormente
6 Infiere la utilidad de la información para tomar de decisiones, recordando evidencia

Fase de prueba ítem de relevancia

<p>¿Qué podrían hacer si algunos niños están contagiados y otros no?</p> <p style="text-align: center;">Indicadores de Relevancia</p> <p>Anticipar que se deben implementar nuevas acciones preventivas, o incrementar el cuidado, o evitar los errores que habrían tenido los niños que se infectaron, porque aún se puede prevenir el contagio de todos.</p> <p>Proponer nuevas acciones de cuidado, como el aislamiento y atención medica de los niños infectados y cambios en las conductas de prevención de quienes no se infectaron</p>

Respuestas de algunos participantes		Cal.
1	Si algunos están bien que se cuiden más y no se pueden quedar en la misma casa	1
2	Se tienen que ir y que cuiden a los enfermos y los otros seguir con su tapabocas	1
3	Si unos están bien se deben separar de los hermanos y no hacer eso, ni acercarse	1
4	Les toca usar tapabocas y lavar el baño	0
5	A todos los tienen que cuidar y si están enfermos mejor se van donde la abuela o otra pieza que esté limpia	1
6	Los que si van al hospital, los otros se quedan pero no pueden salir a la calle, si tosen también, les toca lavar el baño o toda la casa	1

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia
1 Infiere la probabilidad de evitar el evento generando condiciones inhibitoras, usando información de acciones no incluidas en los contenidos de la situación
2 Infiere conductas necesarias para alterar la probabilidad de que repita el evento, usando su conocimiento de condiciones no presentadas
3 Infiere la probabilidad de inhibir el efecto, usando conocimiento de factores inhibidores no presentados (distanciamiento físico)
4 Responde usando la información presentada, descuidando el conjunto de conclusiones establecidas durante la situación
5 Infiere la probabilidad de evitar el evento usando evidencia y adicionando información de las condiciones que mejoran dicha probabilidad
6 Infiere condiciones que aumentan la probabilidad del evento, usando evidencia e información adicional sobre conductas preventivas

Figuras 18 a 26. Dos actividades del IMERIN con explicación de calificaciones.

Propiedades psicométricas del IMERIN

Los análisis estadísticos descriptivos iniciales, confirmaron que los indicadores de acierto en el ítem son confiables para la identificación de conclusiones plausibles y para evaluar juicios generalizantes, explicaciones y predicciones. Por estas razones, los ítems permitieron obtener las calificaciones requeridas para conformar un puntaje en cada dimensión del IMERIN.

Los supuestos que orientaron el análisis estadístico descriptivo de las calificaciones de los ítems ($N = 24$) y de los puntajes de los componentes del instrumento ($N = 96$) fueron: a) altas frecuencias y porcentajes de acierto indican disponibilidad del conocimiento y representaciones requeridas para hacer conclusiones plausibles a partir de los contenidos del ítem; b) porcentajes de acierto más bajos son un primer indicador del nivel de dificultad y discriminación de los ítems; y c) los promedios en generalización son más altos y presentan menor dificultad (probablemente sean más fáciles de responder). (Anexos N, O y P)

La variable observada fue la calificación del ítem (que es categórica) y la escala usada fue dicotómica; se calificó con 1 la coincidencia entre la respuesta y alguno de los 2 indicadores de acierto y con 0 la falta de coincidencia, el puntaje total corresponde con la sumatoria de todos los puntos en los ítems. Las frecuencias y porcentajes de la calificación 1 indican tendencia en el desempeño de los participantes, en las Tablas 13 a 15 se presentan las frecuencias y porcentajes de acierto en cada ítem.

Tabla 13.

Frecuencias y porcentajes de la calificación en cada ítem dominio Ciencias (N=8)

Ciencias	Ítem		Actividad 1 Explicación		Actividad 2 Predicción		
			Frecuencia	%	Ítem	Frecuencia	%
Generalización	1	Calificación 0	6	6.25	5	3	3.12
		Calificación 1	90	93.7		93	96.8
Precisión	2	Calificación 0	14	14.5	6	7	7.29
		Calificación 1	82	85.4		89	92.7
Coherencia	3	Calificación 0	21	21.8	7	23	23.9
		Calificación 1	75	78.1		73	76.0
Relevancia	4	Calificación 0	47	48.9	8	45	46.8
		Calificación 1	49	51.0		51	53.1

Los ítems de generalización tienen los porcentajes de acierto más altos, entre 84.3 % y 97.9%. La frecuencia más alta de acierto en el dominio Ciencias está en el ítem de generalización de la actividad de explicación; esto significa que 6 participantes no hicieron un juicio generalizante plausible y 90 si generalizaron adecuadamente, como se esperaba por tratarse de un ítem representativo de la capacidad de inferir y usar reglas de relación causal, que se ha desarrollado a los 3 años aproximadamente.

Los ítems de las dimensiones precisión y coherencia siguen una tendencia similar en cuanto a tener un porcentaje de acierto mucho mayor que el de falla. En cambio, los ítems 4 y 8 de relevancia en las 2 actividades fueron respondidos de forma acertada por, aproximadamente, la mitad de los participantes. Llama la atención el porcentaje de falla más alto en el ítem 4 de relevancia (48%), en el cual se requería disponer de información relevante sobre la relación entre el cambio temporal de apariencia de las personas y condiciones incidentales, o intenciones no asociadas causalmente con sus necesidades vitales. Este resultado indica, que los ítems de la dimensión relevancia tienden a discriminar mejor entre inferencias basadas en la comprensión de que los cambios, temporales, en la apariencia de las personas son independientes de cambios de apariencia de los insectos, estos últimos asociados con necesidades vitales y adaptación al entorno.

Tabla 14.

Frecuencias y porcentajes de la calificación en cada ítem dominio Estadística (N=8)

Estadística	Ítem	Actividad 1 Explicación		Actividad 2 Predicción			
		Frecuencia	%	Ítem	Frecuencia	%	
Generalización	9	Calificación 0	2	2.08	13	8	8.33
		Calificación 1	94	97.9		88	91.6
Precisión	10	Calificación 0	15	15.6	14	16	16.6
		Calificación 1	81	84.3		80	83.3
Coherencia	11	Calificación 0	46	47.9	15	43	44.7
		Calificación 1	50	52.0		53	55.2
Relevancia	12	Calificación 0	53	55.2	16	52	54.1
		Calificación 1	43	44.7		44	45.8

Respecto a las frecuencias y porcentajes de acierto en el dominio Estadística (Tabla 14) sucedió algo similar a lo descrito sobre los ítems de Ciencias: los aciertos en los ítems de generalización fueron los más altos en ambas actividades, de hecho, el ítem 9 tuvo el porcentaje de acierto más alto de todo el instrumento, 97.9%. Sin embargo, como se explicará en el análisis de consistencia interna del instrumento, calificar el ítem no haría diferencia en la estimación de la capacidad de generalización. También en este dominio se encontraron los porcentajes más bajos de acierto en comparación con los de las demás actividades, en los ítems de relevancia (12) y de coherencia (15), ambos 44.7 %. Frecuencia y porcentaje de acierto de los ítems de relevancia fueron más bajos que en las otras dimensiones (alrededor de 49%). En coherencia los desempeños adecuados fueron cercanos al 50%, que representa igualdad, aproximada, entre las frecuencias de coincidencia de la respuesta con el indicador (aciertos) y de no coincidencia (fallas).

En el dominio Tecnología, la frecuencia más alta de acierto ocurrió en los ítems de generalización, sin embargo, en comparación con los demás dominios, en estos ítems los promedios de acierto fueron más bajos (Tabla 15). Así mismo, se evidencia que la calificación de acierto tiende a presentar promedios ligeramente más bajos en los ítems de las tres dimensiones de la eficacia (rango entre 29.1% y 82.2%). A partir de estos resultados se interpretó la tendencia en los ítems del dominio Tecnología a presentar mayor dificultad y mayor nivel discriminativo de la habilidad (posteriormente se confirmaría esta tendencia).

Las medias del puntaje por dimensión, generalización $M=5.5$; precisión $M=5$; coherencia $M=3.7$ y relevancia $M=2.7$ (máximo puntaje posible 6) evidencian diferencias de magnitud entre las variables estimadas. Dichas diferencias indicaron que los participantes tienen en común la capacidad de usar conocimiento conceptual, es decir, conocimiento de propiedades de las

categorías, de sistemas de causalidad y de covariación. Además, las medias en las dimensiones precisión, coherencia y relevancia mostraron que los puntajes son notablemente diferentes entre ellos y que, específicamente, se distinguen por la relevancia o el sesgo de las representaciones usadas para completar la información del ítem, cuando la evidencia que éste contiene es indirecta en relación con la conclusión solicitada.

Tabla 15.

Frecuencias y porcentajes de la calificación en cada ítem dominio Tecnología (N=8)

Tecnología	Ítem		Actividad 1 Explicación		Actividad 2 Predicción		
			Frecuencia	%	Ítem	Frecuencia	%
Generalización	17	Calificación 0	15	15.6	21	4	4.17
		Calificación 1	81	84.3		92	95.8
Precisión	18	Calificación 0	26	27.0	22	17	17.7
		Calificación 1	70	72.9		79	82.2
Coherencia	19	Calificación 0	56	58.3	23	26	27.8
		Calificación 1	40	41.6		70	72.9
Relevancia	20	Calificación 0	68	70.8	24	42	43.7
		Calificación 1	28	29.1		54	56.2

Validez de constructo y estimación de diferencias individuales en la eficacia del RI.

Los supuestos que orientaron la interpretación de diferencias en los resultados son los siguientes: a) los puntajes de generalización deberían presentar correlaciones bajas con el puntaje total, dado que sus ítems no son necesarios, porque no son componentes críticos de la medición de eficacia, aunque su estimación es un criterio de validez ecológica de la prueba; b) los puntajes de la dimensión y el puntaje total del instrumento pueden ser usados para diferenciar desempeños característicos de los niveles de eficacia en la inferencia; y c) si la confiabilidad es una cualidad psicométrica del instrumento, la interpretación del puntaje se considera válida y útil en el estudio del desarrollo temprano de la habilidad inferencial inductiva.

El puntaje total de la prueba significa la cantidad de aciertos, éste estuvo en un rango de 13 a 24 (máximo posible 24 puntos), con *Media= 17,17* y *Desviación estándar= 3,8* ($N=96$). Mediante coeficientes de correlaciones, biserial puntual y policóricas, se analizaron asociaciones y cambio de puntajes de cada ítem en relación con los dominios, demás ítems y puntaje total de la prueba. Una vez se confirmaron los supuestos sobre confiabilidad y discriminación de los ítems, en relación con diferencias en las conductas observadas, se probó la hipótesis sobre dimensionalidad y estructura del instrumento, por medio del análisis de componentes principales.

Se calcularon los *coeficientes de correlación biserial puntual* (*rpbis*) de los ítems con el puntaje total. El resultado indica que 23 de los 24 ítems discriminan adecuadamente (rango *rpbis* 0,21 a 0,60), o son aceptables en términos de su aporte a la estimación de diferencias de desempeño a partir del total de aciertos (Tabla 16). Esto significa que las puntuaciones discriminan entre las respuestas representativas del atributo (coinciden con el indicador), de las que no lo representan (no coinciden respuesta e indicador).

En el análisis de correlaciones se destacó: que el ítem 18 de precisión del dominio Tecnología es buen predictor de alto puntaje en la prueba completa, (*rpbis*=0.60). Además, se confirmó que la mayor parte de los ítems de generalización son innecesarios para la medición. Incluso el ítem 12 no es aceptable (*rpbis*= 0.00), porque no está correlacionado con el resultado de la prueba, de manera que usarlo no tendría efecto en la estimación requerida. Lo anterior se explica porque los ítems de generalización no tienen indicadores de atributos de la eficacia.

Tabla 16.

Medias, distribución y discriminación de los puntajes en los ítems del IMERIN (N=96)

		Item	Media	Desv.	Rpbis
Ciencias explicación	Generalización	1	0,93	0,2	0.31
	Precisión	2	0,85	0,3	0.58
	Coherencia	3	0,78	0,4	0.56
	Relevancia	4	0,51	0,4	0.26
Ciencias predicción	Generalización	5	0,96	0,1	0.28
	Precisión	6	0,92	0,2	0.27
	Coherencia	7	0,76	0,4	0.49
	Relevancia	8	0,53	0,4	0.34
Estadística explicación	Generalización	9	0,97	0,1	0.0066*
	Precisión	10	0,84	0,3	0.21
	Coherencia	11	0,52	0,4	0.34
	Relevancia	12	0,44	0,4	0.44
Estadística predicción	Generalización	13	0,91	0,2	0.24
	Precisión	14	0,83	0,3	0.39
	Coherencia	15	0,55	0,4	0.42
	Relevancia	16	0,45	0,4	0.49
Tecnología explicación	Generalización	17	0,84	0,3	0.51
	Precisión	18	0,72	0,4	0.60**
	Coherencia	19	0,41	0,4	0.52
	Relevancia	20	0,29	0,4	0.47
Tecnología predicción	Generalización	21	0,95	0,1	0.38
	Precisión	22	0,82	0,3	0.55
	Coherencia	23	0,72	0,4	0.25
	Relevancia	24	0,56	0,4	0.49

Nota: D.E.= Desviación Estándar; *Rpbis*= coeficiente de correlación biserial puntual; *ítem no aceptable; ** ítem más discriminante del instrumento.

Considerando la naturaleza ordinal de la variable aciertos (tienen origen en un indicador), se estudió la consistencia interna del instrumento mediante *coeficientes de correlaciones policóricas* (Elosua y Zumbo, 2008). A partir de los coeficientes se obtuvo evidencia adicional de confiabilidad de la escala de estimación, usando el acierto total en las 4 dimensiones evaluadas. Las correlaciones policóricas fueron: entre generalización y precisión, 0.57; coherencia y generalización, 0.39; relevancia y generalización, 0.39; precisión y coherencia, 0.55; precisión y relevancia, 0.48; y entre coherencia y relevancia de 0.30.

Las correlaciones policóricas entre puntajes en los dominios y niveles de eficacia fueron significativas. Estos datos indican que las puntuaciones de cada dominio se pueden considerar correlacionadas y que es necesario tener puntajes en los tres para obtener estimaciones de eficacia confiables y para establecer niveles del desempeño. Estos resultados concuerdan con lo esperado, teóricamente, para el funcionamiento interno de los componentes del instrumento. Por lo tanto, los coeficientes apoyan el supuesto de homogeneidad de los ítems y justificaron el análisis factorial de la matriz de correlaciones, por el método de componentes principales. Este método se usa para probar hipótesis sobre la dimensionalidad y fiabilidad en la conformación de la estructura de atributos (variables) de la eficacia en la inferencia. En la Figura 27 se presenta el gráfico de las correlaciones entre los dominios y su respectivo coeficiente.

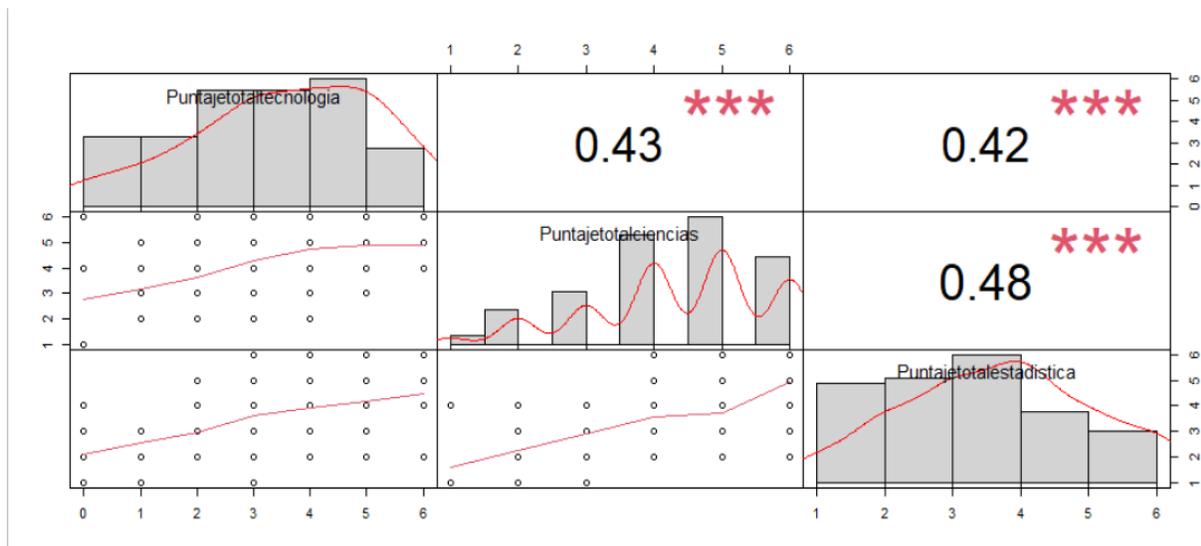


Figura 27. Gráfico de Coeficientes de correlación policórica interdominios. Nota: ** = correlación significativa al 5%.

Mediante el análisis de componentes principales, ACP, se probó la hipótesis sobre unidimensional aproximada de la medida. Este método permitió la identificación de dos componentes en los que contribuyen las 4 variables observadas, el primero explicó el 0.59 de la varianza de los datos y el segundo el 0.17. En el primer componente, generalización contribuye a la covarianza en 0.77, precisión en 0.86, coherencia en 0.72 y relevancia en 0.69; en el segundo componente contribuye coherencia con 0.54, sin embargo, esta contribución es más baja que en el componente 1, por lo que se consideró más significativa en el primero. La Figura 28 presenta las contribuciones de forma gráfica.

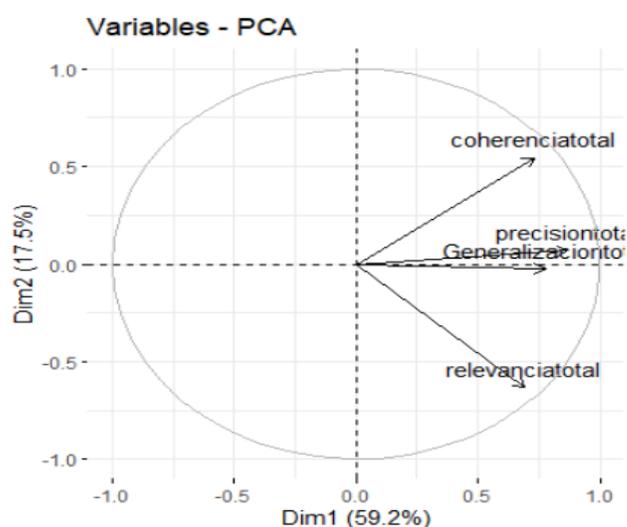


Figura 28. Gráfico de la estructura de componentes principales de la medición de eficacia mediante el IMERIN.

Niveles de eficacia en el RI y patrones de agrupación en las puntuaciones. Con el fin de completar la evidencia sobre validez del instrumento para la medición requerida y siguiendo el plan de análisis, se procedió a comparar la asociación de la puntuación de cada ítem de eficacia con *los rangos* del puntaje total. Considerando que diferencias en la distribución de puntajes de los componentes es evidencia de su utilidad para discriminar niveles de habilidad inferencial, el análisis se realizó mediante la prueba estadística Kruskal -Wallis (prueba K).

Mediante la prueba Kruskal-Wallis (estadístico H , prueba K) se estudió la contribución individual de los ítems ($N=18$) que conforman cada dimensión de eficacia a la varianza total del puntaje. Esta es una prueba no-paramétrica alternativa al análisis de varianza cuando no se cumplen supuestos, en particular el supuesto de distribución normal. Las medias son: Precisión 5,

Coherencia 3.7 y Relevancia 2.7, con el nivel de significancia $p \leq 0.010$, se estableció que todos los ítems contribuyen significativamente a la distribución de varianza en la subescala en la que se inscriben. En otras palabras, todos los ítems están midiendo nivel de eficacia en la inferencia, de manera que cuando no se acierta, la estimación del atributo cambia de magnitud. Los p -valores del estadístico H se presentan en la Tabla 17 (Anexo L). (explicación de la *prueba K* en Canavos, 1999, pg 598)

Tabla 17.

Valores p del estadístico H-prueba K para las subescalas de medición del IMERIN.

	Subescala	Item	p valor H
Ciencias explicación	Precisión	1	7.155e-07*
	Coherencia	2	1.245e-07*
	Relevancia	3	0.0077
Ciencias predicción	Precisión	1	0.0303
	Coherencia	2	2.98e-06*
	Relevancia	3	0.0006
Estadística explicación	Precisión	1	0.0146
	Coherencia	2	0.0012
	Relevancia	3	3.824e-06*
Estadística predicción	Precisión	1	0.0006
	Coherencia	2	0.0001
	Relevancia	3	2.545e-07*
Tecnología explicación	Precisión	1	1.558e-08*
	Coherencia	2	1.858e-07*
	Relevancia	3	6.148e-07*
Tecnología predicción	Precisión	1	1.103e-06*
	Coherencia	2	0.0270
	Relevancia	3	3.962e-06*

Nota:*=notación científica, indica cantidad de ceros (0) a la izquierda del dato. Nivel de significancia $p \leq 0.010$

Las calificaciones de los ítems se agruparon con referencia a la mediana en las subescalas de eficacia (rango entre 0 y 6), con el fin de comparar niveles de desempeño. Se observaron las siguientes tendencias generales: a) a usar conocimiento de propiedades diversas en las categorías, en los tres dominios, independientemente de que éstas resultaran novedosas (podrían ser desconocidas para el participante hasta el momento de la familiarización); b) a evaluar la evidencia directa, para usarla como guía en la inferencia de nuevos casos de la relación o categoría razonada; y c) una menor disponibilidad de información relevante para guiar la inferencia cuando la evidencia de nuevas relaciones es indirecta, en los tres dominios. Los niveles de calificación se observan en la Figura 29.

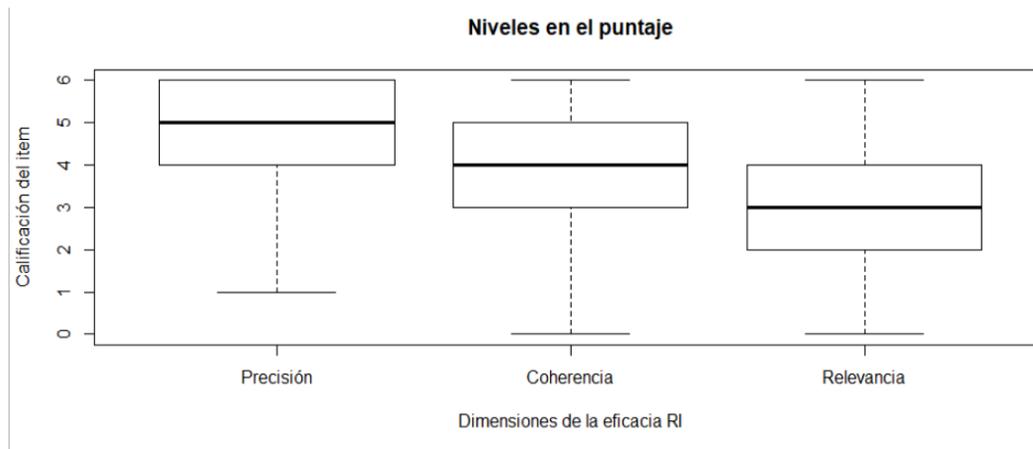


Figura 29. Gráfico de niveles de puntuación en las dimensiones de eficacia del RI. La línea en el centro de las cajas representa la Mediana: Mediana precisión 5, Mediana coherencia 4, Mediana relevancia 3, $N=96$.

Como se mencionó, la creación de contenidos del IMERIN siguió los estándares propuestos por AERA, APA y NCME, (2014) en relación con los criterios que se deben cumplir para que la evaluación de habilidades cognitivas sea justa (pg. 50). Por esta razón se estudió la influencia de las variables individuales edad y género en las diferencias del puntaje total y en cada subescala o atributo de la eficacia del RI. Acorde con lo esperado, los resultados indican que ninguna de las puntuaciones de los ítems y de las subescalas se asocia significativamente con edad, a nivel de significancia del 1% ($p \leq 0.010$), valores *rho* entre .0 y 0.1. Las medias del puntaje de las subescalas en cada subgrupo de edad se presentan en la Tabla 18.

Estos resultados proveen evidencia empírica que apoya la hipótesis de investigación desde la perspectiva del desarrollo, sobre la relativa estabilidad de los subprocesos cognitivos implicados en la eficacia del RI durante la edad preescolar. Probablemente, en esta etapa de la vida, el cambio conceptual es más significativo para la eficacia del RI que las estrategias de inferencia. Este cambio ocurre a partir de la transformación del conocimiento de dominio específico, especialmente, por la comprensión progresiva de sistemas de causalidad y de estructuras de las categorías (Hayes, 2007). Lo anterior coincide con la falta de asociación entre los puntajes de las subescalas de eficacia y las edades de los participantes (p valores entre 0.1 y 0.9; *rho* entre 0.01 y 0.1) y con el hallazgo de diferencias significativas en los puntajes del dominio Tecnología en función del género del participante.

Tabla 18.
Medias de los puntajes en subescalas del IMERIN en cada subgrupo de edades.

Edad	n	Media Pun. total	Media Precisión	Media Coherencia	Media Relevancia
4	46	17.6	5	3.5	2.9
5	43	17.6	5	3.9	2.5
6	7	20.9	5.4	4.4	4

Nota: *M*= media. Puntaje total incluye aciertos en generalización.

Interpretación de diferencias en los niveles de eficacia entre los participantes. De acuerdo con los datos de la Tabla 18, las diferencias de puntuación en la dimensión relevancia, entre el grupo de participantes con edades entre 4 y 5.11 años y el de 6 años, es el aspecto que caracteriza el desempeño y determina su clasificación en algún nivel de eficacia. Aunque las diferencias en los puntajes no son estadísticamente significativas, los datos muestran que hay una leve ventaja de los participantes mayores de 5 años, en los puntajes de Precisión, en términos generales estos se clasificaron en el nivel alto en todos los grupos. Respecto a la trayectoria en el desarrollo del RI eficaz, parece haber dos momentos: entre los 4 y 5. 11 años es similar el nivel de habilidad inferencial y después de los 5.11 años este nivel mejora.

La relación entre edad y nivel de eficacia podría relacionarse con una mejoría progresiva de la habilidad para evaluar evidencias a partir de las teorías previas, como lo sugiere Deanna Kuhn (1990, 2005, 2010b). La diferencia en los puntajes de los ítems de Coherencia y Relevancia de los niños mayores, indica que los participantes usaron múltiples fuentes de información y conocimiento para interpretar la evidencia directa e indirecta. Como se explicará, más adelante, la disponibilidad de diversas fuentes de información, especialmente, sobre la existencia de propiedades causales invisibles, y sobre la relación entre las tecnologías y la agencia humana, puede ser el factor clave en las diferencias individuales en el desempeño eficaz de las actividades del IMERIN.

Usando la prueba Wilcoxon de diferencia de medias y medianas (rango del puntaje entre 3 y 6) se halló diferencia estadísticamente significativa entre los puntajes promedios de los niños y de las niñas en el dominio Tecnología, $p=0.004$ (nivel de significancia del 0.05; Ciencias $p=0.3$ y Estadística $p=0.7$, no significativos). Esta diferencia probablemente también se asocie con estereotipos de género, que marcan diferencias en las oportunidades que tienen las niñas de acceder, usar y apropiarse de Tecnologías, en especial en los países en vía de desarrollo (OECD, 2015). Las

medias de puntaje en cada subescala para el grupo de niñas y el de niños se presentan en la Tabla 19.

Tabla 19.

Medias de los puntajes en subescalas del IMERIN discriminados por género.

Genero	<i>N</i>	<i>M</i> Puntaje total	<i>DE</i>	<i>M</i> Precisión	<i>M</i> Coherencia	<i>M</i> Relevancia
Niñas	39	16,53	3,7	5	3.8	2.8
Niños	57	17,6	3,8	5	3.6	2.7
Total	96	17,17	3,8	5	3.7	2.7

Nota: *M*= media, *DE*= desviación estándar. Puntaje total incluye aciertos en generalización.

En conjunto, estos resultados confirman el supuesto sobre independencia entre atributos medidos por la prueba. Así mismo, los datos son indicio del adecuado control del error de medición y representan evidencia de que las estimaciones de atributos de la eficacia son homogéneas, e independientes entre ellas.

Estudio 2c. Acuerdo interevaluadoras: Validez de criterio y validez predictiva del índice de habilidad inferencial

Con el objetivo de estudiar elementos de validez externa y de criterio, y para favorecer la replicabilidad de la indagación del RI mediante el IMERIN, 3 evaluadoras repitieron la calificación de los ítems de las subescalas de eficacia con el 30% de las aplicaciones ($N= 26$). Los resultados de la segunda evaluación permitieron: a) establecer un criterio empírico para determinar si hubo errores en la primera medición, atribuibles a sesgos de la evaluadora en la interpretación de los indicadores de puntuación; b) probar la objetividad atribuida a los indicadores, mediante el análisis de los acuerdos entre observadoras en la asignación de puntaje en cada ítem; y c) obtener evidencia adicional de la capacidad de discriminación de perfiles de habilidad de las subescalas de eficacia, mediante el análisis de correlaciones (*Rho*, Spermán) entre los primeros y segundos puntajes.

Plan de análisis e interpretación. Además de los análisis estadísticos y psicométricos de la segunda calificación de la prueba, se realizó un estudio detallado de los conceptos y sugerencias de las evaluadoras en relación con las estrategias de documentación de la experiencia y los diversos ámbitos de utilidad de los índices de habilidad inferencial derivados de la estimación. La información se usó para ajustar el manual de aplicación, delimitar supuestos sobre la utilidad

diagnostica del perfil de RI obtenido mediante el IMERIN y considerar otros usos posibles del mismo, en diversos ámbitos de la Educación y la Psicología.

En promedio, las calificaciones de las tres evaluadoras coinciden 98% con las de la primera, en los 26 casos estudiados. El coeficiente de correlación entre las 6 puntuaciones del primer grupo fue $rho=.97$; para las 10 del segundo, $rho=.99$; y las 10 del tercer grupo, $rho=.98$. Estos resultados fueron interpretados como evidencia de la alta coincidencia en el uso del indicador de puntuación de la mayoría de los ítems, en las dos evaluaciones realizadas. Se confirmó que los indicadores son suficientemente discriminativos de las conclusiones precisas, coherentes y relevantes, por lo que es posible afirmar que permiten la objetividad en la evaluación del desempeño en la prueba, independientemente del criterio del evaluador o del posible sesgo asociado a su campo de experticia.

Las observaciones registradas por las evaluadoras en los formatos de la segunda calificación indican que las fallas en el audio dificultaron la comprensión de las respuestas, razón por la cual se podría explicar la ausencia de calificación en algunos de los ítems o la falta de coincidencia entre las dos puntuaciones (7 ítems diferentes, en total). Esta información se usó para plantear algunas limitaciones en ambas modalidades de la aplicación, que deberán considerarse en futuros estudios del RI que hagan uso del IMERIN, o de instrumentos similares.

Las evaluadoras entrevistadas consideran que las diferencias en los índices de habilidad no se asocian al género del participante, aunque en ciertos dominios, por ejemplo, Biología (niñas) y Tecnología (niños) las conclusiones de niñas, o de niños son más amplias y las referencias a las categorías son más precisas. Es probable que las diferencias se asocien con preferencias por razonar sobre ciertos contenidos, es decir, con creencias estereotipadas. En el dominio Biología se podrían asociar, además, con creencias sesgadas respecto a la clase ideal de miembros de la categoría (Foster-Hanson et al, 2020).

Desempeño en los dominios STEM y tendencias en el desarrollo de habilidades inferenciales.

Para obtener evidencia adicional de la confiabilidad del puntaje para el diagnóstico cognitivo, se analizaron particularidades distintivas del desempeño de la prueba en relación con el contenido de los indicadores de eficacia. Para este propósito se usó el método estadístico llamado Arbol de decisión, el cual genera algoritmos para separar los datos en grupos distintivos de acuerdo

con reglas de clasificación que el mismo modelo establece¹. De esta forma, el modelo clasifica los patrones de coincidencias entre puntuación, dominios y variables individuales, del contexto y de la situación de aplicación. Para el análisis se ingresaron todos los datos disponibles, sin embargo, el modelo generó reglas usando únicamente variables del desempeño. Es decir, los demás datos resultaron irrelevantes para crear los criterios de separación, agrupación y nuevas clasificaciones.

En la estructura de clasificación, o árbol de decisión, el primer rectángulo representa la regla usada para separar los datos en dos categorías, de desempeño en este caso. Los 2 rectángulos ubicados en el segundo nivel de la estructura representan nuevos criterios de agrupación de los datos que conforman cada una de las dos categorías, los demás rectángulos también representan criterios de agrupación, para los datos a los que no aplicó la regla de clasificación inmediatamente anterior. En los cuadrados ubicados al final de la rama se muestra el número de observaciones en cada grupo y el porcentaje de la muestra que fue clasificado. Las estructuras de clasificación (árboles de decisión) de los resultados de la aplicación del IMERIN se presentan en las Figuras 30 a 32.

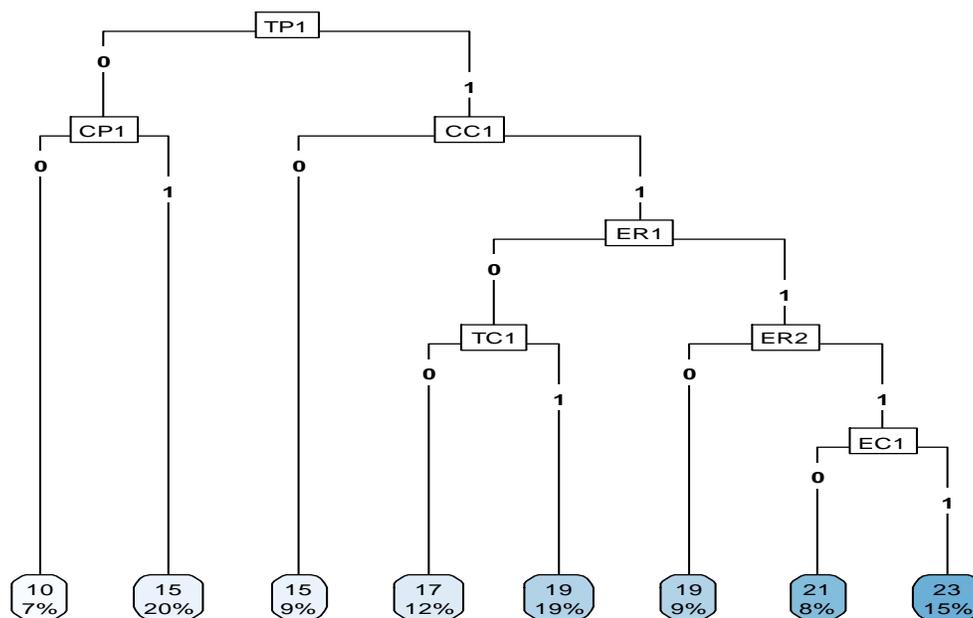


Figura 30. Estructura de clasificación de patrones distintivos del desempeño en la fase de prueba del IMERIN, o “Árbol de decisión”. Nota: T es Tecnología, C es Ciencia y E es Estadística; P ítem de precisión, C (segundo lugar) coherencia y R relevancia; 1 corresponde a la modalidad explicación y 2 a predicción.

¹Descripción del método Árbol de decisión en: <https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>

Los resultados de los análisis de clasificaciones de las respuestas permitieron: a) identificar patrones distintivos del RI de los participantes con habilidades inferenciales; b) confirmar las particularidades del desempeño en los 3 niveles de eficacia, a partir de los indicadores usados para la clasificación; c) emplear el criterio de separación inicial de los datos y los posteriores de agrupación, para predecir la pertenencia a una clase o nivel de habilidad según el dominio; y d) comparar las reglas de clasificación con los criterios para la segunda evaluación del IMERIN. En conjunto, los resultados sustentan la confiabilidad de la interpretación de características distintivas del desempeño, que son rasgos del perfil de habilidades inferenciales en los dominios. Además, los hallazgos de la clasificación permiten precisar recomendaciones para el uso diagnóstico de los puntajes de eficacia en el RI.

Conforme a la clasificación presentada en las estructuras de los árboles de decisión, el ítem que mejor discrimina entre los niveles de habilidad inferencial es del dominio Tecnología, modalidad explicación y evalúa precisión, TP1 en adelante. La separación de los datos a partir de la calificación en TP1 implica que las características distintivas de los perfiles de habilidad están asociadas a la inferencia de múltiples causas no obvias, o de propiedades causales invisibles interactuando en un sistema, para producir el efecto observado. El ítem de Coherencia en la modalidad explicación del dominio Ciencia, CC1, sigue en importancia para la agrupación de patrones de desempeño independientes.

El acierto en el ítem de precisión de esta actividad, TP1, hizo la diferencia entre las puntuaciones de nivel medio (17 a 20) y alto (21 a 24), en los grupos de niños de 4 y 5 años (Tabla 19). La situación en la que se enmarca este ítem hace explícito el sistema de relaciones y los mecanismos de causalidad probables para el efecto magnetismo. De acuerdo con el análisis de dificultad y discriminación, realizado en la primera parte del estudio de resultados, la inferencia que evalúa este ítem resultó más compleja que las demás en el IMERIN. En TP1 se presenta evidencia directa de un nuevo caso en el que ocurrió el efecto, pero ésta es indirecta en relación con la intervención del agente que aseguró la contigüidad, o contacto entre imanes y objetos metálicos (Bonawitz et al., 2019; McCormark, 2016), las acciones del agente se observaron en la fase 1. El análisis de esta actividad y las respuestas de 6 participantes se presenta más adelante (Figuras 33 a 36).

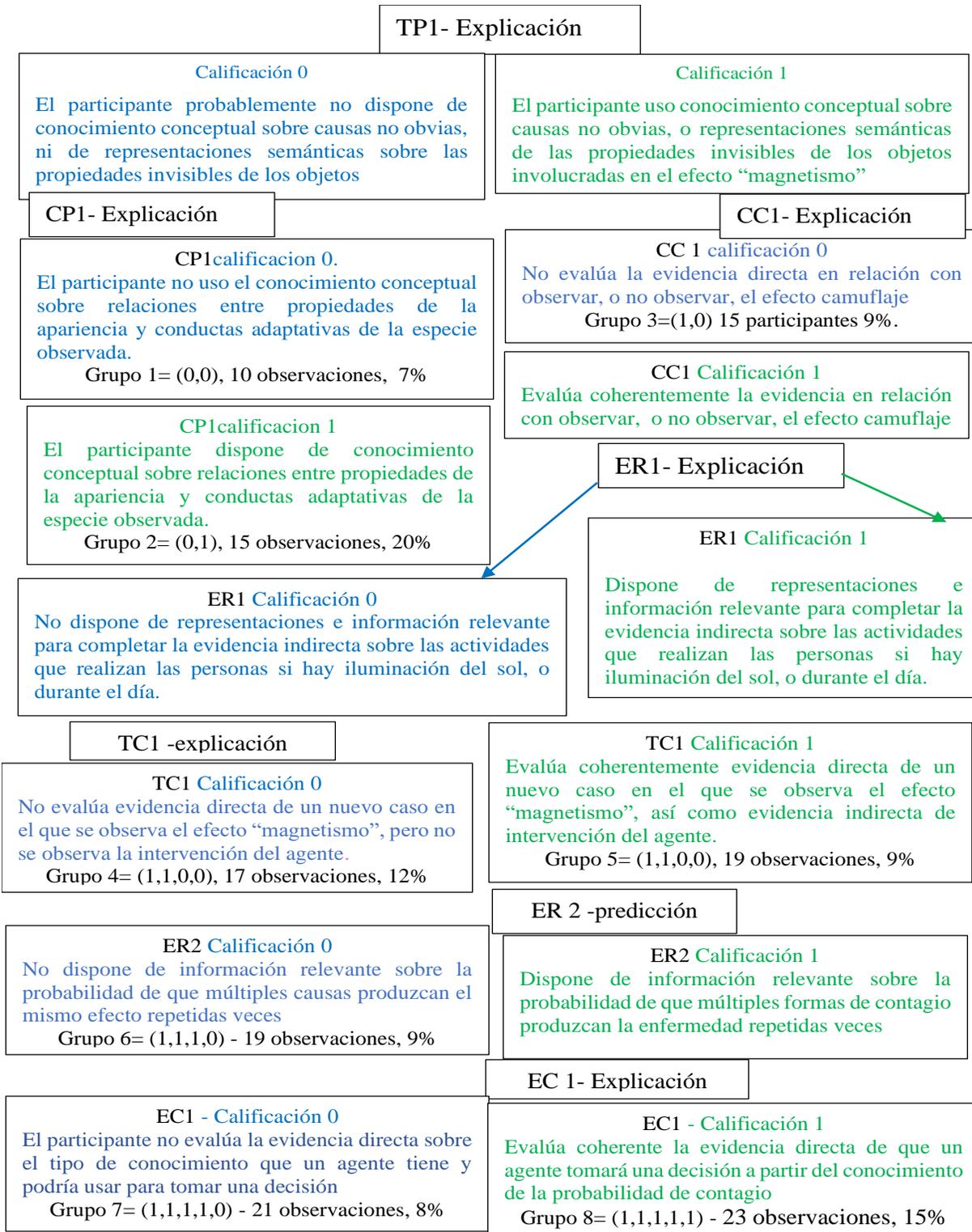


Figura 31. Clasificación de patrones de desempeño en el Arbol de decisión con especificación de indicadores de puntuación de los ítems significativos.

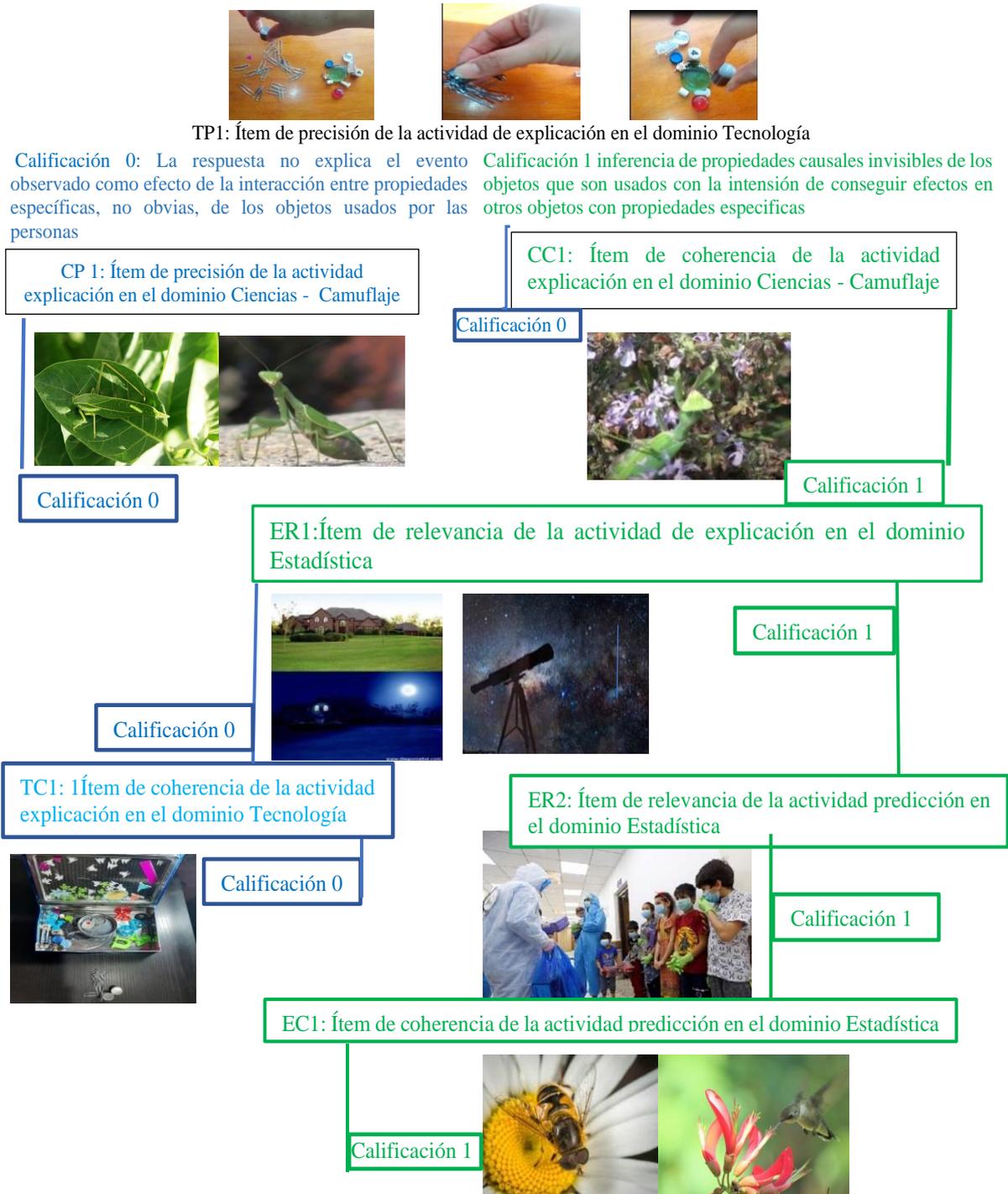


Figura 32. Imágenes - evidencia en los ítems de la estructura de clasificación de patrones distintivos del desempeño en el IMERIN, o “Árbol de decisión”.

Actividad Tecnología – Explicación

Objetivo: inferir que, propiedades no obvias como el “magnetismo” y el “ferromagnetismo”, interactúan si se interviene para asegurar que los objetos que las poseen se acerquen o se mantengan juntos. La conclusión eficaz depende de considerar la diversidad de contextos y muestras, así como diversas propiedades invisibles de los objetos y artefactos.

Contexto: el RI eficaz requiere conocimiento de la probabilidad de ver efectos sin tener evidencia de las causas. Además, de experiencias previas en las que la intervención es la condición para observar interacciones y efectos de esas propiedades (usar el control remoto del TV., por ejemplo).

Fase inicial familiarización: inicia con la presentación de un video en el que se observa la intervención de un agente que acerca imanes a diferentes clases de objetos, unos metalicos y otros de diferentes materiales. Luego se informan los nombres de los objetos y de las interacciones presentados

¿Para qué se acerca el imán a los diferentes objetos?




Objetivo ítem: generalizar la regla o condición acercamiento para generar el efecto magnético; esta es acercar el imán a objetos metálicos; sin la acción del agente el efecto no ocurre

Indicadores de la plausibilidad de la generalización

- Explica que se estaba demostrando que el imán atrae los clips y estos se quedan pegados, y que hay algunos objetos que no son como los clips y por eso no se pegan al imán
- Asocia la acción de un agente que acerca el imán, con la intención de probar si los objetos poseen características no obvias por las cuales se queden pegados
- Usa evidencia de acercamiento para discriminar cuáles objetos tienen propiedades no visibles de categoría causal y cuales no

Respuestas de algunos participantes		Cal.
1	Para que se peguen los clips	1
2	Porque tiene imanes	1
3	Quiere pegarlas al imán y porque son de botón no se pegan	1
4	Quiere jugar con las fichitas de clips	1
5	Porque los quiere pegar	1
6	Quiere que se peguen con el imán	1

Justificación de la calificación del ítem a partir del uso de los indicadores de plausibilidad
1 Generaliza la condición necesaria para observar el efecto presentado
2 Generaliza las propiedades causales, no obvias, que intervinieron en el evento
3 Generaliza su conocimiento de las propiedades distintivas de las categorías
4 Generaliza la intención de usar las propiedades de las categorías para generar un efecto
5 Generaliza la intencionalidad del agente como condición del efecto visto
6 Generaliza la intervención del agente como parte de las condiciones del sistema causal

Figuras 33. Actividad de explicación Ítem de generalización, del dominio Tecnología, con interpretación de calificaciones de 6 participantes.

Fase de prueba ítem de precisión	
	
<p>¿Porque solo se pegaron los clips al imán y los otros objetos no?</p> <p style="text-align: center;">Indicadores de Precisión</p> <p>Explicar que los clips están hechos de metal y los otros objetos de otros tipos de material</p> <p>Afirmar que hay diferentes tipos de materiales o composiciones entre los objetos, y los clips son de la clase de materiales que se pegan a los imanes</p>	

Respuestas de algunos participantes		Cal.
1	Porque el imán pega el metal y el plástico no se pega	1
2	Porque son de metal, no se los otros que son	1
3	Porque los clips tienen metal, las fichitas no	1
4	Esos se pegan y los otros no	0
5	Así son de pegarse	0
6	Porque tienen metal	1

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia
1 Infiere las propiedades causales usando conocimiento preciso sobre las categorías
2 Infiere algunas propiedades causales, expresando su conocimiento de las categorías
3 Infiere las propiedades exactas de las categorías con eficacia en el efecto y sin eficacia
4 Refiere las imágenes de las situaciones
5 Repite respuestas omitiendo alguna información presentada durante la actividad
6 Infiere parte de las causas intervinientes en el efecto indagado

Figuras 34. Actividad de explicación Ítem de precisión, del dominio Tecnología, TP1, con interpretación de calificaciones de 6 participantes.

La coincidencia de la conclusión con el indicador de puntuación depende de usar conocimiento sobre la existencia de causas invisibles, propiedades no obvias y diversidad en el origen de las relaciones causales (Zimmerman, 2000). Es decir, la conclusión depende de inferir procesos de causalidad a partir de sus resultados (Koloswky, 1996; Walker et al., 2014). La inferencia precisa (calificación 1) indica el descubrimiento de la interacción entre causas suficientes y causas necesarias, pero no suficientes; además, da cuenta de la proyección de propiedades diversas y no obvias del dominio de la Física a otros dominios (Bluter, 2012).

Fase de prueba ítem de coherencia



Introducción: Al abrir la caja metálica, algunos de los diferentes objetos que estaban adentro se quedan pegados a la tapa, ¿porqué pasa esto?

Indicadores de Coherencia

Explicar que los objetos que permanecen pegados tienen imanes que no se ven, en el interior o están hechos con imanes

Suponer que se observa el efecto de la acción de un agente que metió -acercó objetos con imanes a la caja, con el objetivo de que se quedaran pegados a la tapa.

	Respuestas de algunos participantes	Cal.
1	Debe ser que tienen imán de pegarse por detrás	1
2	No se sabe	0
3	Porque son de imán, tu tapas y se pegan	1
4	Porque están entre la cajita	0
5	Tienen pegamento?	0
6	Porque se guardaron y después se abrió y ahí se pegaron	1

Justificación de la calificación del ítems a partir del uso de los indicadores de eficacia

1 Infiere propiedades no visibles que tienen eficacia en el evento razonado
 2 Ignora la evidencia y no usa las conclusiones e información anteriores
 3 Infiere las causas y condiciones necesarias para el efecto usando evidencia implícita en la situación presente e información de sus anteriores conclusiones
 4 y 5 La información de las respuestas no son indicadores de inferencia
 6 Infiere la intervención como condición necesaria para que ocurra el evento usando evidencia directa

Figuras 35. Actividad de explicación Ítem de coherencia, del dominio Tecnología, TC1, con interpretación de calificaciones de 6 participantes.

En el ítem de coherencia TC1 se evalúa si la inferencia integra de forma coherente el juicio sobre el valor informativo de la evidencia con la conclusión del ítem inmediatamente anterior. Si dicho juicio es coherente con la nueva evidencia indirecta, posiblemente el razonador reevaluó la probabilidad de cambio en las propiedades causales, o en la relación causal generalizada previamente (Griffiths et al., 2012). Este ítem tiene la particularidad de que involucra la expectativa del participante frente a la probabilidad de observar el efecto, así que la evidencia directa debe usarse para confirmar o desconfirmar lo anticipado. En este orden de ideas, la conclusión coherente

es evidencia de que los modelos mentales del participante guían razonamientos adaptativos (Nisbett y Ross, 1986).

Fase de prueba ítem de relevancia	
	
<p>¿Por qué las figuritas que adornan la puerta de la nevera se quedaron pegadas ahí?</p> <p style="text-align: center;">Indicadores de Relevancia</p> <p>Aclarar que se dio la intervención de alguna persona que aproximó los adornos a la nevera, porque sabe que éstos contienen imanes y que el efecto de pegarse se mantiene hasta una nueva intervención para separarlos.</p> <p>Afirmar que la nevera es de metal y solo acercando los adornos estos se pegan y permanecen unidos sin más intervención</p>	

Respuestas de algunos participantes		Cal.
1	Serán de imanes	0
2	Un humano las puso ahí para hacer más bonita su nevera	1
3	Que tal si alguien colocó las canastas y esas mariposas con imán	1
4	Porque ellas volaron	0
5	Porque son de metal	0
6	Para que se vean	0

Justificación de la calificación del ítem a partir del uso de los indicadores de eficacia

- 1 La respuesta se refiere a las características de las categorías
- 2 Infiere condiciones y causas necesarias y suficientes del evento indagado, a partir de evidencia implícita y haciendo uso del conocimiento de las circunstancias contextuales que generaron la relación
- 3 Infiere las causas necesarias para el efecto indagado, haciendo uso de sus creencias
- 4, 5 y 6 las respuestas refieren conclusiones anteriores no usan información adicional

Figuras 36. Ítem de relevancia de la actividad explicación, en el dominio Tecnología, TR1, con interpretación de calificaciones de 6 participantes.

El desempeño en el ítem de relevancia de esta actividad fue determinante en el nivel alto de eficacia, en todas las edades (Tabla 18). La frecuencia de acierto fue 28 ($N=96$), lo cual indica que la adición de información relevante referida a la evidencia indirecta de la intervención (condición necesaria para el efecto), resultó demandante para los niños menores de 6 años. Además, las conclusiones improbables expresaron creencias irrelevantes en relación con la agencia humana.

En especial, sobre la capacidad de intervenir resultados, haciendo uso del conocimiento de propiedades no obvias y diversas de los objetos del sistema de causalidad razonado. En conjunto, los resultados de la dimensión relevancia indicarían que, al parecer, el conocimiento previo de la capacidad de apropiar recursos materiales y tecnológicos es insuficiente antes de los 6 años.

Conclusión preliminar. El análisis de resultados de los estudios de validez predictiva apoya los supuestos respecto a la confiabilidad del uso diagnóstico del puntaje en el IMERIN. Puesto que se establecieron adecuadas propiedades psicométricas de instrumento, acuerdo entre las evaluadoras independientes y coincidencia entre sus conceptos y las puntuaciones usadas para la clasificación del árbol de decisión. Además, los análisis estadísticos y cualitativos de las diferentes evaluaciones, determinaron que patrones del desempeño individual estaban en el mismo nivel de habilidad (básico, medio y alto) en el que fue ubicado el puntaje en la primera evaluación. En síntesis, la coincidencia en las 2 interpretaciones de características significativas en el patrón de desempeños, aporta evidencia de validez externa del resultado de la medición realizada mediante el instrumento.

Los datos y evidencias analizados indican que las estrategias metodológicas permitieron el control sistemático de diversas fuentes de error en la medición. Como son: la operacionalización del constructo, la calidad y presentación de los contenidos, los sesgos de los evaluadores, la modalidad de aplicación y la influencia de variables de la situación de observación, o contextuales no relevantes. Por lo tanto, resultó oportuno continuar con el estudio de los perfiles de habilidad inferencial característicos en el grupo de participantes.

Perfil característico del RI eficaz de alto nivel

Las características particulares del RI eficaz, se interpretan como índices de habilidades inferenciales de los participantes. En esta parte del análisis de perfiles de los participantes se incluirán las puntuaciones ubicadas en el nivel alto de eficacia. Las características mencionadas son las siguientes:

a) Disponibilidad de conocimiento e información relevante relacionada con el uso y apropiación de propiedades causales invisibles, o internas, de ciertas categorías asociadas con las tecnologías. Los niños con edades entre 4 y 5. 11 años, referían en sus conclusiones información del ese tipo de propiedades, con una frecuencia similar, en aproximadamente el 50 % de los casos. Los participantes mayores, con edades entre 5.11 hasta 6 años, lo hicieron en el 75% de los casos.

b) Generalización de evidencia directa e indirecta, a partir de la inferencia de relaciones de causalidad probabilísticas.

c) Inferencia de propiedades de la agencia humana y de resultados de su intervención en los procesos de cambio, o en las circunstancias que delimitan un resultado. Incluso, se infiere que no hubo influencia de la intervención humana en un sistema de relaciones, o en los cambios en las propiedades internas y de la apariencia de las categorías, a partir de la evidencia.

d) Proyección de la evidencia a nuevos casos de la relación inferida, incluso adicionando información relevante sobre circunstancias y condiciones diversas de los contextos, que tendrían eficacia en los cambios anticipados.

e) Explicación de razones por las cuales no es posible afirmar, o concluir relaciones en los casos en los que no hay evidencia de cambios o covariaciones. Es decir, se identifican en la respuesta, nociones sobre azar y probabilidad condicional que apoyan inferencias sobre ausencia de cambios, ausencia de intervención o ausencia de evidencia que justifique alguna conclusión sobre relaciones pasadas o futuras.

El perfil de RI de los participantes, cuyos puntajes los ubican en el nivel medio de eficacia, se caracteriza por las siguientes capacidades:

- a) Un aspecto común en las respuestas, que es crítico en la demostración de la habilidad inferencial, es que éstas expresan que el participante comprendió la necesidad de hacer explícitas sus representaciones semánticas de la evidencia indirecta. Probablemente, este tipo de respuesta tiene como finalidad compartir, con el interlocutor, la información que usó como guía del razonamiento.
- b) Los participantes difieren notablemente en la profundidad de la explicación inferida. Algunas explicaciones apenas coinciden con el indicador y otras se caracterizan por la adición de piezas de información, ilustrando experiencias significativas con los contenidos de la situación razonada.
- c) En los casos en los que el participante no tenía experiencia con situaciones de razonamiento similares y la evidencia resultó novedosa, ésta se usó como referente para hacer hipótesis sobre relaciones similares a las inferidas anteriormente. También se evidenció agrado (incluso preferencia) por descubrir, o adivinar, cuáles eran las variables no explícitas asociadas al efecto, relación o evento razonado.

- d) Algunos participantes con puntuaciones levemente superiores a la media, informaron sobre las claves sociales o pedagógicas que usaban como referente para superar la incertidumbre frente a los contenidos. Es probable, que dichas claves se relacionaran con diferencias de eficacia en función del dominio de la actividad.

En general, el puntaje del grupo muestra que no hubo desempeños en el nivel bajo. Puntajes individuales por debajo de 16 (de 24), se consideraron en el nivel básico de eficacia. Estos puntajes podrían indicar algunas dificultades para inferir propiedades invisibles, o creencias sesgadas que influyen en la atención y valoración de la evidencia directa. Así mismo, desconexión entre la generalización inicial y las inferencias posteriores. Esto a su vez, probablemente se asocie con inseguridad, o falta de confianza frente al conocimiento disponible para razonar sobre las categorías y relaciones en la actividad. El significado de la falta de coincidencia entre las respuestas del participante y el indicador, fue interpretada por las 3 evaluadoras (estudio de validez y confiabilidad) de forma similar.

Estudio 3. Resultados de la aplicación de la encuesta

La estrategia de encuestar a las madres de los niños y niñas participantes surgió en respuesta a la necesidad de generar evidencia, ecológicamente válida y útil en el análisis del supuesto sobre la influencia de las actividades de mediación del desarrollo conceptual de los menores *habituales o frecuentes*. Los resultados indican que las categorías de prácticas de socialización son fuentes de datos relevantes sobre los hábitos de comunicación del conocimiento, experiencias de búsqueda de evidencias y demostraciones de relaciones consideradas relevantes para el aprendizaje de conceptos. También muestran, que la frecuencia probablemente es una magnitud adecuada para indagar la influencia de las prácticas socializantes en el desarrollo del RI eficaz.

Plan de análisis e interpretación de resultados

El primer análisis de las respuestas al cuestionario se centró en establecer cuales prácticas de socialización fueron habituales para la mayoría de los participantes (o un porcentaje amplio de ellos) y determinar si compartían características generales en el grupo. Esta información se analizó mediante técnicas estadísticas descriptivas, a partir de las especificaciones de objetivos y patrones de conductas que definen las categorías de prácticas estudiadas. Inicialmente se analizó el patrón

de hábitos, para esto se obtuvo la cantidad de participantes que eligieron la opción correspondiente a la frecuencia alta en cada ítem. Cuando hubo empates entre los hábitos, para ordenar la jerarquía, se tomó en cuenta la cantidad de participantes que reportaron frecuencia media en estos mismos. De esta manera también se estableció el orden de preferencias de los participantes por cada clase de prácticas.

Los ítems que indagaron opinión y preferencias por áreas del saber se analizaron en conjunto con el ítem que preguntaba por la percepción de igualdad en las oportunidades de acceder al conocimiento para ambos géneros. Para este estudio se realizó el análisis de correspondencias simples y se describieron características de las actividades que se agruparon a partir de aspectos comunes en el patrón de coincidencia de las respuestas.

El estudio de relaciones entre hábitos y puntaje del instrumento se realizó a partir del coeficiente de correlación Spearman. El estadístico confirmó que el puntaje total en el IMERIN no está correlacionado con alguna práctica (ni habitual, ni esporádica). Este resultado se encontró coherente con la expectativa de resultados de esta etapa del estudio. Puesto que, la evidencia obtenida en el estudio 2, sugiere que las diferencias en puntajes de eficacia del RI se explican por diferencias en las habilidades inferenciales. Por esta razón, se concluyó que la influencia de otros tipos de diferencias individuales, o de variables contextuales, deberá ser objetivo de futuras investigaciones del desarrollo conceptual.

También se estudió la asociación entre regularidades de las prácticas de socialización de la cognición y diferencias en los promedios de eficacia en cada dimensión y dominio, es decir, se compararon las varianzas y las medias de los *conjuntos* de datos. Para realizar este análisis se usó la prueba estadística Kruskal-Wallis (estadístico H). Los resultados se usaron para identificar tendencias en las actividades de mediación del desarrollo conceptual y hábitos de formación de la cognición que pueden considerarse, relevantes, para la indagación de factores sociales que influyen en el desarrollo del RI de los niños colombianos.

Por último, mediante la técnica de arboles de decisión se analizaron los criterios de clasificación de las relaciones entre puntajes medios del IMERIN y frecuencias en las prácticas de socialización. Estas relaciones se muestran gráficamente, en una estructura que representa la separación de los grupos de observaciones que estaban más alejadas entre sí y luego, muestra los pares asociados. Al modelo de clasificación se vincularon las variables sociodemográficas y

contextuales consideradas de importancia para describir el grupo de participantes y las condiciones contextuales que influyeron en el proceso de la encuesta.

Estudio 3^a. Regularidades en las actividades familiares de formación del pensamiento de los niños

La evidencia estadística y la información documentada sobre las respuestas a la encuesta, indican que algunas de las prácticas habituales correspondían con *metas* concretas para el aprendizaje y adquisición de conocimientos de los menores. Estas metas serían guías en la mediación familiar del desarrollo cognitivo de los niños. De otra parte, las actividades de búsqueda de conocimiento y pedagógicas, que se realizaban esporádicamente, se orientaron por *expectativas* relacionadas con afrontar consecuencias del cierre de las aulas. Estas se concretaron en experiencias didácticas sugeridas por profesoras y acciones de acompañamiento a los niños en el proceso de adaptarse a la dinámica de enseñanza-aprendizaje a distancia. En la Tabla 20 se resumen datos descriptivos de las prácticas habituales en el grupo de participantes.

Tabla 20.

Prácticas de socialización de la cognición habituales entre los participantes

#	Ítem	N - f3
1	Cuando hablamos, uso los nombres y palabras exactas para mencionar los objetos, hechos, personajes o sujetos a los que me refiero	55
2	He aprendido sobre los temas o situaciones que le generan curiosidad, preguntas o interés.	42
3	He confirmado que la información que cree que es verdadera, realmente si lo es, por ejemplo: información sobre cómo evitar enfermedades causadas por virus y bacterias	41
4	Propongo conversaciones que le faciliten aprender el significado de palabras relacionadas con temas de su interés.	38
5	Doy ejemplos para explicarle que las personas tenemos la capacidad de pensar y generar conocimientos, desde que somos niñas y niños	38
6	Le cuento sobre los nuevos aprendizajes que yo he logrado y le pregunto sobre los suyos.	33

Nota: N - f3 es la cantidad de participantes que eligieron la frecuencia 3, o alta.

El hábito más frecuente y común entre los participantes es usar los nombres y palabras exactas para mencionar los objetos, hechos, personajes o sujetos en sus conversaciones. La práctica habitual de esta actividad comunicativa, además de hacerse con el objetivo de que los niños

aprendan a hablar y a usar el lenguaje adecuadamente, busca brindar claves lingüísticas que faciliten la expresión del razonamiento y favorezcan el desarrollo de conocimientos sobre las categorías de importancia social (Butler & Tomasello, 2016). Es probable que la intención comunicativa subyacente a este hábito sea la de compartir conocimiento de propiedades determinantes de las categorías, aquel que se considera adecuado enseñar y aprender para el desarrollo temprano de habilidades de solución de problemas (Cerchiaro & Puche, 2018).

El *segundo hábito* común entre las participantes es aprender sobre los temas o situaciones que le generan curiosidad, preguntas o interés al menor. La actividad se relaciona con verificar la confiabilidad de la información que se usa para instruir sobre conceptos, para crear situaciones pedagógicas centradas en transmitir conocimientos a los que los niños han atribuido importancia y en los que las madres están dispuestas a invertir recursos personales. La frecuencia de esta práctica es indicio de la aceptación que el área o conocimiento específico genera en la familia, probablemente las inquietudes de los niños sobre el tema son compartidas por las madres, de manera que la búsqueda de información confiable, en fuentes externas, se considera la estrategia adecuada para resolverlas.

El hábito de confirmar que la información que el niño cree que es verdadera, realmente si lo es, resultó en *tercer* lugar entre las prácticas que la mayoría de los participantes realizan con alta frecuencia. La particularidad de este resultado radica en que esta práctica es la única de la categoría B que había sido habitual durante el mes anterior a la encuesta. La categoría B se refiere a involucrar al niño en un patrón de actividad con el fin de que conozca, se enfoque, o preste atención a ciertas regularidades en las relaciones observables, o en cierta clase de evidencia. Por lo tanto, dada la relevancia social que adquirió el tema de las enfermedades causadas por virus (desde el 2020) y el hecho de que ninguna otra práctica de la categoría tuvo alta frecuencia entre la mayoría de los participantes, es probable que su inclusión entre los hábitos relevantes sea consecuencia de la familiaridad con la clase de conceptos incluidos en el ejemplo que se usó en el ítem.

De acuerdo con la suposición anterior, en la interpretación del resultado se tuvo en cuenta que la frecuencia reportada en esta práctica podría ser representativa de un patrón de actividades de socialización de la cognición característico del contexto de crisis sanitaria. Más adelante en este capítulo se discuten algunas implicaciones, para el desarrollo conceptual de los niños y para la educación Preescolar, de la baja representación de la categoría B entre los hábitos de formación del pensamiento de los niños.

Alrededor del 50% de las participantes habitualmente proponían conversaciones que facilitarían al menor aprender el significado de palabras relacionadas con temas de su interés. Esta práctica comunicativa implica que se atiende a las áreas de conocimiento que interesan a los niños y se usa el lenguaje para el aprendizaje conceptual en las mismas. Probablemente, la conversación sea usada por ambos como estrategia principal para transmitir significados en cualquier campo. La información adicionada a las respuestas en este ítem, indica que los niños también usan la conversación para aclarar los significados de palabras relacionadas con los temas que les interesan. De manera que este hábito vincula aspectos de la socialización cognitiva diferentes a los considerados en la creación de la encuesta, como el fomento al desarrollo de la autonomía del niño.

Dar ejemplos a los niños para explicar la capacidad de pensamiento de todas las personas, desde la infancia, tiende a ser una práctica habitual para la mayoría de los participantes. La inquietud que originó el ítem se relaciona con obtener información sobre la posibilidad de que en las familias se piense en la igualdad de género como propósito de formación del pensamiento. De hecho, dar ejemplos a los niños podría ser la actividad que involucre la capacidad de las madres para dar significado a las experiencias y aprendizajes cotidianos. Kuhn, (2010) se refiere a la capacidad de dar sentido a la experiencia, como un mecanismo significativo (e implícito) en la construcción del pensamiento científico en la infancia. Por estas razones, la práctica habitual de demostraciones de la capacidad cognitiva, mediante el ejemplo, puede considerarse un factor importante en la implementación del enfoque de la educación STEM en el Preescolar.

Por otra parte, la categoría a la que pertenece el hábito en referencia trata aspectos relativos a la reflexión sobre la influencia de la comunidad en la valoración personal del conocimiento y en la elección de estrategias de trasmisión y mediación para el desarrollo de éste. Además de reportar altas frecuencias de práctica, el resultado de la encuesta indica que el ítem fue de interés para las participantes, quienes lo mencionaron durante la entrevista.

El enunciado del último ítem que tuvo frecuencia significativamente alta es el siguiente: *“Le cuento sobre los nuevos aprendizajes que yo he logrado y le pregunto sobre los suyos”*. Esta fue una práctica menos habitual, pero con cierta periodicidad y fue común entre un porcentaje superior al 50% de los participantes, de manera que fue poco frecuente, o esporádica entre el otro 50%, sin llegar a ser significativamente desconocida entre las madres que no la realizaban. Pertenece a la categoría de comunicación del conocimiento y es una estrategia de seguimiento directo al desarrollo de conocimientos, que implicaría para la madre actualizarse, si esperaba estar

al tanto de los nuevos aprendizajes del niño y brindarle retroalimentación adecuada. Es importante mencionar que cuando esta práctica *no fue frecuente*, las puntuaciones medias en las dimensiones y dominios del IMERIN, tendieron a ser más altas, especialmente en Tecnología.

En síntesis, las actividades practicadas con regularidad constituyen patrones de comunicación que dirigen atención e intereses del niño hacia cierta clase de información, o de evidencia a la que se atribuye importancia en la construcción del conocimiento. Dichos patrones también son procesos, socialmente encauzados, que se caracterizan por estar dirigidos intencionalmente. Probablemente, con el fin de acumular conocimientos, cambiar y formar nuevas representaciones conceptuales en dominios específicos (Ricco, 2015). De manera que el vínculo entre regularidades en las actividades, intenciones, creencias y estrategias de comunicación del conocimiento influye el desarrollo conceptual de forma directa y, por ende, en la eficacia de la inferencia indirectamente.

Estudio 3b. Prácticas de socialización de la cognición que influyen el puntaje del IMERIN

El objetivo de esta etapa del estudio, fue encontrar relaciones directas e indirectas entre las metas y expectativas de desarrollo conceptual de las madres y las estimaciones de eficacia y nivel de desempeño de los menores en situaciones de RI. El análisis de correlaciones (Spearman) entre el puntaje total en el IMERIN y las frecuencias de las actividades de socialización corroboró que datos de frecuencias, desagrupados, no se relacionan directamente con las diferencias en el RI individual. Este resultado apoya la hipótesis sobre la relación de las prácticas de socialización, habituales, con el desarrollo y cambio conceptual, que implicaría la influencia indirecta en la precisión, coherencia y relevancia de la inferencia. De hecho, se confirmó la asociación de dos hábitos y dos prácticas en las diferencias de puntajes promedio en algunas subescalas del instrumento (ver Tabla 23). En el Anexo M se presentan los coeficientes de correlación *Rho*.

Se usó la prueba Kruskal-Wallis para examinar relaciones entre resultados de la encuesta y estimaciones realizadas con el IMERIN. Los valores *p* significativos ($p \leq .010$) muestran asociaciones entre las medias del puntaje en 1 dimensión de la eficacia y 1 dominio de las actividades de RI, en los 4 niveles de frecuencia en la práctica. En la Tabla 21 se presenta la información de las práctica, dimensión y dominio asociados, con los respectivos *p*-valores.

Tabla 21.

Asociaciones estadísticamente significativas entre prácticas de socialización de la cognición y puntajes del IMERIN.

Ítem	IMERIN	P-valor
7 Realizamos actividades para descubrir las causas de eventos que podrían generarle curiosidad, por ejemplo: mezclar dos colores para obtener un tercer color diferente	Relevancia	*0.005
	Ciencias	*0.015

Nota: p - valor obtenido por medio de la prueba Kruskal-Wallis. ** valor significativo.

La práctica de descubrir causas no obvias que intervienen en los efectos, o cambios físicos que sorprenderían o generaban curiosidad a los niños, fue un hábito común entre un porcentaje importante de las participantes (40%) y resultó asociado con los puntajes promedio en el dominio Ciencias y la dimensión Relevancia. Los datos asociados muestran una relación de direccionalidad inversa; al aumentar la frecuencia de la práctica disminuye el puntaje promedio en la dimensión, o dominio correspondiente al conjunto de respuestas en ese nivel de frecuencia. En general, la tendencia en la asociación indica que probablemente las nociones, o información sobre causas no obvias, comunicadas a los niños por las madres y otros miembros de la familia, no influyeron en la relevancia de la información que los niños usaron para razonar sobre contenidos del dominio de la Física y la agencia humana.

Otro hábito que mostró tendencia a cambiar de frecuencia según el nivel de desempeño en el IMERIN (primero en la Tabla 22), corresponde con el siguiente enunciado: “*le cuento sobre los nuevos aprendizajes que yo he logrado y le pregunto sobre los suyos*”. El aumento en la frecuencia de esta actividad se relacionó con diferencias en los puntajes de eficacia en la dimensión Precisión y en los dominios Tecnología y Ciencias. Además, la práctica frecuente de esta actividad ocurrió entre los participantes que estuvieron vinculados a alguna institución de cuidado o de Educación Preescolar durante el año 2019.

La relación de direccionalidad entre niveles de frecuencia de esta actividad (ítem 14) y medias en precisión y relevancia, indica que probablemente, esta práctica afecta la pertinencia y suficiencia de las representaciones semánticas y conceptuales que los niños consideran fuente de información relevante para razonar, y que usan para hacer explícitas sus conclusiones. Esta suposición se relaciona con evidencia de que los niños adquieren patrones de conocimiento de la realidad a través del discurso del adulto (Mercer, 2013)

Tabla 22.

Frecuencias de las prácticas de socialización de la cognición habituales y los puntajes promedio en subescalas de eficacia del IMERIN correspondientes.

Habito	N	Media Precisión	Media Coherencia	Media Relevancia	Media Punt.tota
1					
Nunca	0	-	-	-	-
Poca	4	6	4.5	4.2	21
Media	22	4.9	3.8	2.9	18
Alta	55	4.9	3.8	2.6	17.5
2					
Nunca	1	6	4	2	19
Poca	7	5	3.6	3.1	17.7
Media	31	5.2	4	2.7	18.3
Alta	42	4.7	3.7	2,7	17.5
3					
Nunca	3	6	4	4	20
Poca	6	5	3.8	2.2	16.8
Media	31	4.8	3.6	2.8	17.4
Alta	41	5	3.9	2.7	18.1
4					
Nunca	1	2	2	1	10
Poca	14	5.1	4	2.8	18
Media	28	4.9	3.6	2.8	17.6
Alta	38	5.1	3.9	2.8	18.1
5					
Nunca	1	2	2	1	10
Poca	14	5.1	4	2.8	18
Media	28	4.9	3.6	2.8	17.6
Alta	38	5.1	3.9	2.8	18.1
6					
Nunca	1	6	4	0	17
Poca	13	5.2	4	3.1	18.6
Media	29	4.9	3.5	2.8	17.3
Alta	38	4.9	4	2.7	17.9

Por último, se observó la tendencia a aumentar las frecuencias en la actividad de lectura de temas deseables para el desarrollo conceptual del niño (ítem 15), si disminuyen los puntajes promedios en Ciencias. Dado que la mayor parte de las participantes reportaron intereses en enseñar sobre áreas del conocimiento diferentes a las STEM, es probable que los temas de los textos leídos a los niños, no fueran relacionados con los contenidos del IMERIN.

Un hecho importante que podría haber influido en la tendencia de la asociación reportada es la desescolarización y restricción al contacto físico y social durante el 2020, porque ambas condiciones limitan el acceso a oportunidades para nuevos aprendizajes, los cuales dinamizan el desarrollo cognitivo. Incluso, la crisis social también podría haber afectado drásticamente el

involucramiento de las madres (y otros familiares) en la elección y acceso a recursos culturales, pedagógicos y tecnológicos para apoyar prácticas de socialización de la cognición.

Estudio 3c. Diferencias en la eficacia del RI y en las prácticas de socialización de la cognición asociadas a variables individuales.

Como se describió anteriormente, se usó la prueba Kruskal-Wallis, para estudiar relaciones entre diferencias de puntaje en el IMERIN, frecuencias de las prácticas estudiadas y las variables: edad de los niños, género, vínculo institucional previo, número de hermanos, empleo y nivel de escolaridad de la madre. Adicionalmente, se usó la técnica árboles de decisión para comparar asociaciones entre las variables referidas y los valores promedio en la frecuencia de actividades habituales y esporádicas de socialización de la cognición. Las Figuras 37 y 38 presentan los árboles de decisión.

Los resultados indican que, algunas de las prácticas que se relacionaron con diferencias de puntuaciones promedio en las dimensiones de eficacia del RI, también se asociaron con el hecho de haber estado vinculado a una institución de Preescolar durante el año 2019 y con la cantidad de hermanos. Este resultado señala la importancia de disponer de recursos para la mediación del desarrollo cognitivo de los niños. Tales como, el tiempo y acceso a la orientación por parte de expertos en servicios de educación o de cuidados en la infancia temprana. De hecho, la única práctica común que tenían las madres de niños que no tuvieron vínculo institucional antes de 2020, fue inventar historias poco creíbles sobre hechos conocidos, para motivar a sus hijos a verificar la exactitud del relato (h11). La mayor parte de ellas la realizaron con poca frecuencia.

Las madres de los niños que estaban vinculados a una institución durante el 2019 y que tenían dos hijos, regularmente usaban nombres y palabras precisas en sus conversaciones con los niños (h12 en el árbol, Figura 37). Además, esporádicamente inventaron actividades para tener la oportunidad de comparar similitudes y diferencias entre objetos y seres presentes en el entorno (h2). Así mismo, la práctica de conversar sobre los nuevos aprendizajes (h14 en el árbol) fue esporádica en este grupo. Dado que la actividad se clasifica en la categoría A, referida a actividades comunicativas que expresan conceptos, preferencias y creencias, el resultado indica, que en los casos en los que los niños no fueron escolarizados, previo a la Pandemia y tenían un hermano, hubo baja frecuencia de diálogos sobre el conocimiento valorado socialmente.

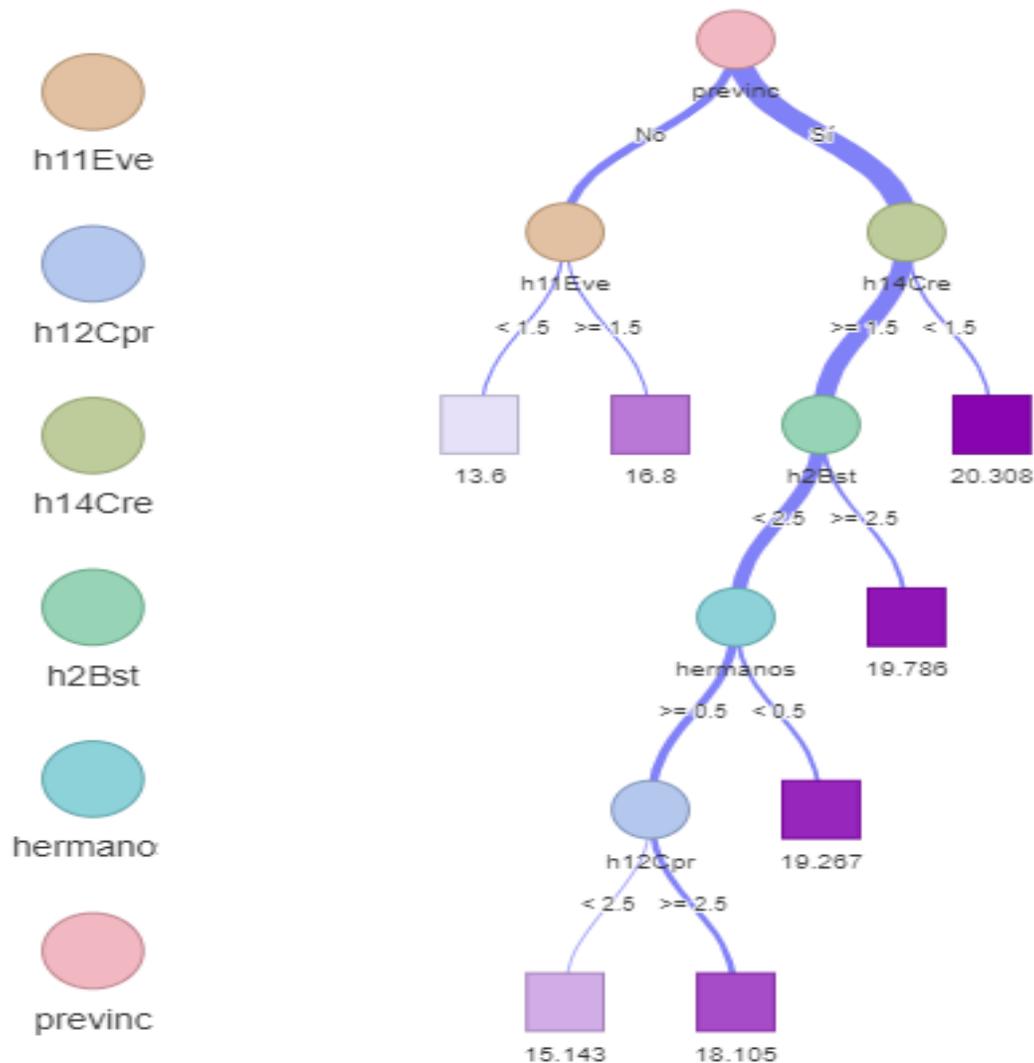


Figura 37. En el árbol de decisión los círculos representan a las prácticas y variables individuales que el modelo incluyó en el análisis de asociaciones. El número en la nomenclatura corresponde con el ítem, las letras se usaron para distinguir cada ítem en las tablas de datos. Los cuadrados representan conjuntos de observaciones en cada frecuencia de la práctica. Los valores bajo el cuadrado son el puntaje promedio del IMERIN que coincide con las observaciones.

En el grupo de participantes que tuvieron vínculo institucional previo (*previnc*), los puntajes promedio en el IMERIN tendieron a aumentar si habían sido hijos únicos. Probablemente, este resultado se asocie con el hecho de que, los recursos para el acompañamiento del proceso de aprendizaje y mediación del desarrollo cognitivo, se concentran en un solo hijo si es el menor, o si no tiene hermanos (Hernández, 2010).

De otra parte, es importante mencionar que el nivel de escolaridad de las madres participantes no mostró relación con los promedios de puntajes de eficacia en RI, ni con los de

frecuencia de las prácticas de socialización de la cognición. Esta variable se podría haber asociado con disponibilidad de recursos para el acompañamiento, como algunas de ellas lo reportaron, espontáneamente. Sin embargo, para afirmar que hay influencia de esta asociación en el desarrollo conceptual de los niños se requiere evidencia directa (OECD, 2014).

También se generó un Arbol de decisión que solo incluye los ítems de la encuesta, sin asociaciones con variables individuales. Este gráfico (Figura 38) muestra la división de las respuestas en dos grupos, estimada en función de la frecuencia de las conversaciones sobre los nuevos aprendizajes entre madres e hijos (h14). Una vez separados los datos de frecuencia regular y baja de esta actividad, aparecen nuevas agrupaciones entre prácticas de socialización y puntajes promedio en el IMERIN, las cuales no se habían encontrado mediante los demás análisis de datos derivados de la encuesta.

Las dos actividades que aparecen agrupadas en el segundo árbol de decisión son: a) Dedicar tiempo para verificar la plausibilidad de los conceptos y teorías que los niños habían aprendido recientemente (h10); y b) realizar actividades para descubrir las causas no obvias de cambios físicos visibles (h7). A su vez, las dos prácticas se asociaron en los casos que ambas presentaron cierta regularidad y si además conversaban sobre nuevos aprendizajes frecuentemente (h14).

En el árbol, la primera agrupación del gráfico muestra que la verificación frecuente de la eficacia de los nuevos aprendizajes (h10) se relaciona con baja frecuencia en la actividad de inventar historias poco creíbles para motivar a los niños a verificar los hechos (h11). En la práctica h11, hubo un porcentaje importante de participantes que nunca la realizaron, o que la realizaron pocas veces durante el mes anterior a la encuesta. Quizás esta actividad fue considerada más compleja que las otras, dado que implica involucramiento de la madre en la revisión de teorías ingenuas de los niños y el fomento al RI en condiciones de incertidumbre. La baja frecuencia en este ítem se asoció con puntajes promedio en el IMERIN por debajo de la media general.

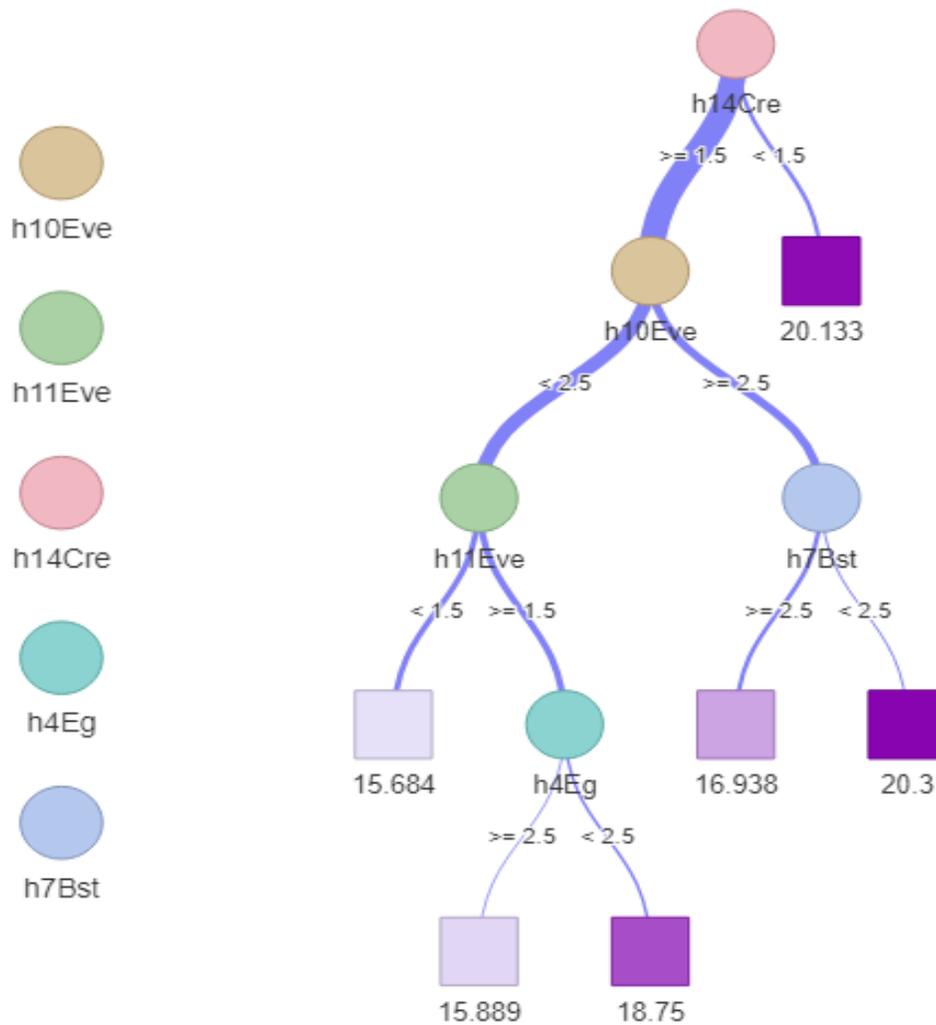


Figura 38. En el árbol de decisión los círculos representan ítems que el modelo estadístico agrupó. El número en la nomenclatura indica cuál es el ítem, las letras se usaron para distinguir enunciados. Los valores bajo el cuadrado son el puntaje total promedio del conjunto en el IMERIN.

En la segunda rama del árbol se asocian los casos (25%) en los que tiende a ser habitual que las madres inventen historias para motivar a su hija o hijo a verificar los hechos (h11) y la búsqueda frecuente, de aprendizaje sobre el conocimiento que generaba curiosidad o interés a los menores (h4). El ítem 4 fue el último asociado a este conjunto de prácticas y es un hábito compartido por una parte importante de los participantes (70% aproximadamente). La relación de esta práctica con el puntaje promedio en el IMERIN sugiere, que los temas que las madres aprendieron, por interés en el desarrollo conceptual de sus hijos, y sobre los que creaban historias

con incertidumbre para motivarlos a verificar información, fueron de áreas distintas a los dominios STEM.

El tercer conjunto de prácticas que se agrupan en el árbol se caracteriza por cierta regularidad en las actividades de descubrimiento de las causas por las que ocurren algunos cambios físicos (h7). Wilkening y Cacchione, (2010) sugieren que las explicaciones de los mecanismos, reglas de causalidad o principios de la Física intuitiva influyen el cambio conceptual durante la infancia temprana. Desde esta perspectiva, y a partir del análisis de los resultados, surgió la hipótesis sobre la confianza de las madres en representaciones conceptuales sesgadas, o irrelevantes para el aprendizaje en el dominio de la Física, la Química y otras Ciencias.

Ejemplo de tendencias distintivas en las prácticas de formación de la cognición. A continuación, se describen tendencias en las prácticas de socialización de la cognición en la familia de un niño que obtuvo el máximo puntaje en el IMERIN. Sus respuestas representan todas las características de RI eficaz, este perfil de habilidades inferenciales se describió en la sección anterior. Se incluyó en la descripción información brindada por la madre durante una entrevista informal. Las tendencias referidas, son las siguientes:

-Habitualmente realizan todas las actividades de comunicación, búsqueda de evidencias y experiencias de descubrimiento de relaciones causales y de nuevo conocimiento sobre la realidad. En su familia es importante prestar atención a los cambios en las condiciones del entorno cotidiano y en las dinámicas sociales actuales.

-Involucran a todos los miembros de la familia (en convivencia) en la formación del pensamiento del niño, manteniendo una perspectiva definida de cómo pueden contribuir ellos con sus objetivos de aprendizaje del conocimiento deseable, el cual corresponde, en este grupo particular con conocimiento de profesiones de áreas STEM y con conocimientos de los oficios apreciados por las madres, como los culinarios y de la repostería, por ejemplo.

-La madre del niño considera muy importante fomentar la orientación al logro por mérito propio, porque considera que la autonomía es la habilidad cognitiva más importante para el desarrollo intelectual de sus hijos. Este principio de la socialización de la cognición fue referido (en la sección de adición de información de la encuesta) por las madres de los 8 niños con puntajes del IMERIN en el nivel superior, entre 22 y 24.

Discusión y conclusiones

Hallazgos y proyección de resultados

La instrumentalización del proceso de indagación de las inferencias eficaces de un grupo de niños habitantes de Bogotá, con edades entre 4 y 6 años, permitió la creación de una metodología válida y confiable para caracterizar el RI temprano. Así, mediante actividades de razonamiento con contenidos de los dominios STEM, fue posible generar nueva evidencia empírica que sugiere que los niños, desde los 4 años, están en capacidad de: generalizar, usar conocimiento conceptual preciso, evaluar coherentemente evidencia directa e indirecta y elegir información relevante para reducir la incertidumbre, usando claves sociales, lingüísticas y pedagógicas. En consecuencia, se propone que la medición de atributos de la eficacia en la inferencia, mediante el IMERIN, permite obtener un índice de la habilidad inferencial individual. El cual se estudiará en profundidad, una vez reestablecida la normalidad en la Educación inicial.

Al probar la validez del IMERIN se avanzó en la creación de herramientas adecuadas para el estudio científico de la influencia del conocimiento conceptual y de regularidades estadísticas en la eficacia del RI de los niños. Este es un hallazgo clave, que podría ser replicado en miras de consolidar aportes al cuerpo de conocimiento de modelos probabilísticos del razonamiento. Los cuales refiere Gopnik, (2011) como aquellos modelos que admiten la incertidumbre del razonador en la medición y formalización matemática de las conclusiones producto del proceso inferencial inductivo. En conjunto, los hallazgos reportados y las nuevas evidencias generadas avanzan en la construcción de unidad teórica y metodológica para el trabajo investigativo en el campo de las teorías ingenuas (Schulz 2012; Gopnik, 2012; van der Graaf et al., 2016).

Se logró el objetivo de crear y proponer herramientas válidas para caracterizar el RI de los participantes, a partir de particularidades de sus inferencias explícitas. Diferenciándolas de otros procesos, como percepción, asociación y abstracción (Rhodes y Gelman, 2008). En ese sentido, la conceptualización de los procesos de razonamiento hizo énfasis en distinguir la inferencia inductiva de los mecanismos subyacentes al CI. Así mismo, en consolidar argumentos a favor del distanciamiento metodológico de los paradigmas de la deducción e inteligencia fluida. En otras palabras, para caracterizar al RI en la infancia temprana, éste debe estudiarse a partir de diferencias psicológicas, singularidades en la inferencia y conclusiones explícitas, dado que éstas son

respuestas observables (Nisbett y Ross, 1980; Sternberg y Prezt, 2005). Evitando de esta forma, la confusión del RI con el razonamiento deductivo y el abstracto.

Desde lo conceptual, este logro dependió de proponer al aprendizaje estadístico implícito como mecanismo subyacente a la inferencia en situaciones de razonamiento con estructura inductiva. Para la orientación metodológica se consideraron, entre otros, tres tipos de antecedentes empíricos: a) estudios de sensibilidad de los bebés a frecuencias en la interacción social cotidiana y a patrones de información estadística, explícita e implícita, reportados por Baillargeon y cols. (2016); b) los estudios de Pavlidou y Bogaerts, (2019), basados en el análisis de evaluaciones repetidas de cualidades del desempeño, como la facilidad (fluidez) y precisión; y c) la indagación del reconocimiento de distribuciones espaciotemporales, considerando la sensibilidad de los bebés a patrones estadísticos visuales, examinados por Tummeltshammer y Kirkham, (2013) y Wu et al, (2011), entre otros.

En ese orden de ideas, es importante resaltar respecto al trabajo metodológico de esta investigación, que el control de la demanda cognitiva, de la suficiencia y relevancia de la evidencia directa e indirecta fue producto del *trabajo multidisciplinario* para la creación e integración sistemática de contenidos a la estructura de la situación de RI. Específicamente, en la fase de familiarización se confirma disponibilidad del conocimiento de nombres y denominaciones de los elementos o categorías y se evalúa la plausibilidad de la generalización de relaciones o principios implicados. Dado que, se halló evidencia de la validez y confiabilidad de la interpretación de resultados del IMERIN, se puede afirmar que las estrategias para el control de la demanda de memoria de trabajo, o de otro tipo de funciones ejecutivas, también fueron efectivas (Lawson y Fisher, 2012).

La perspectiva integradora, asumida para avanzar en la caracterización del RI en la infancia temprana y el trabajo multidisciplinario en los procesos de validación, a pesar de las condiciones limitantes impuestas por la crisis sanitaria, tiene implicaciones para el avance teórico y metodológico de los modelos de desarrollo cognitivo y la educación científica en la infancia. Un antecedente que ejemplifica la necesidad de trabajar en dicho avance se encontró en algunos estudios con niños en edades menores a 5 años. Aquellos, que han probado el desarrollo temprano de la capacidad de hacer inferencias causales a partir de conocimiento de patrones estadísticos y de probabilidades. En estos, se ha asegurado que las situaciones de razonamiento presentaran evidencia *suficiente* sobre la eficacia de los agentes causales. También, se verificó la relevancia del

conocimiento previo sobre propiedades de las categorías implicadas (Gopnik et al, 2001; Sobel y Legare, 2014; Sutherland y Cimpian, 2017; Waldmann y Hagmayer, 2006).

Sin embargo, es común que en los reportes de investigación no se especifique el método usado para determinar el rango o medida en los que la evidencia puede ser suficiente. Tampoco se especifica cómo aseguraron que no se excedió el contenido de información explícita, hasta llegar a transformar la demanda cognitiva, o la naturaleza de la situación y, por ende, del atributo o desempeño observado. En ese contexto, diferentes perspectivas del razonamiento y del desarrollo del pensamiento científico en la infancia, coinciden en que la predicción adecuada de la probabilidad de que, múltiples causas afecten un resultado, mejora con el conocimiento de *evidencia empírica* sobre covarianza (Kuhn et al, 2008; Sternberg, 2010). De igual forma, con el conocimiento teórico (que implica cierto grado de comprensión teórica) de las estructuras causales en dominios y contextos específicos (Kuhn y Dean, 2004).

Lo anterior significa que, los razonadores que disponen de conocimiento conceptual preciso, harán predicciones más eficientes sobre relaciones en sistemas causales, en condiciones de incertidumbre (Sobel et al, 2017). En ese sentido, surgió la pregunta por la influencia del desarrollo conceptual en la eficacia del RI en la infancia temprana. Los hallazgos reportados en la presente disertación avanzan en la búsqueda de respuestas pertinentes a dicho interrogante. Sin embargo, es de esperar que este aspecto de la formación del pensamiento, origine gran parte de las cuestiones a ser abordadas en la investigación de las habilidades inferenciales, deseables, para la vida durante el siglo XXI (Greiff et al., 2014; Griffiths et al., 2012)

Para la indagación de la cuestión planteada anteriormente, se tomó como referente el tipo de trabajos que desarrollaron Samarapungavan y cols. (2008). Ellos crearon prácticas del aula y enfoques didácticos en los planes de estudio del Preescolar a partir de teorías populares, propias en la cultura, referidas a fenómenos astronómicos. Ese conjunto de intervenciones pedagógicas se implementó con el fin de hacer seguimiento al cambio en los patrones de desarrollo conceptual de los niños que participaron y aprendieron los contenidos. De esta forma se encontró y documentó un importante volumen de evidencia de que preferencias, expectativas y deseabilidad del conocimiento están influidas críticamente por la cultura, como lo plantearon Nisbett y cols., (2001).

A su vez, la preferencia de los formadores por algún tipo de conocimiento, deseable para el aprendizaje durante la infancia, se asocia con cambios en el desarrollo conceptual. Por esta vía, la mediación del adulto afecta el aprendizaje de habilidades inferenciales de los niños. De hecho,

mediante la aplicación del IMERIN se comprobó que las conclusiones inferidas inductivamente están orientadas por las teorías intuitivas de los niños y por las creencias sobre las Ciencias, que se favorecen en las diferentes comunidades (Evans y Over, 2013; Nisbett et al, 2001). En síntesis, es razonable considerar que estos resultados aportan conocimiento relevante para la investigación empírica del vínculo entre desarrollo del conocimiento conceptual y aprendizaje de habilidades inferenciales subyacentes al pensamiento científico.

Las pruebas de diferentes tipos de validez del IMERIN, proporcionaron evidencia estadística concluyente que afirma los planteamientos de las teorías de cambio y desarrollo conceptual en la infancia, específicamente en la edad preescolar. Estas conclusiones, apoyan la idea de transferir conocimiento de la Ciencia cognitiva a la Educación. De forma similar lo plantean las conclusiones de trabajos de Carey (2009); Waldmann y Hagmayer, (2006) y Sloutsky y Fisher, (2004), entre otros. En ese sentido, se puede afirmar que, conocer referentes conceptuales y metodológicos válidos y confiables, brinda la oportunidad de transformar la enseñanza, para fomentar el *cambio conceptual*. Este cambio ha sido planteado por investigadores de la Psicología Educativa; Kuhn, (2010a); Mohlhenrich et al., (2018); por divulgadores científicos (OECD, 2009) y por el Ministerio de Educación Nacional, como meta en la implementación del sistema de calidad de la Educación en el siglo XXI, (MEN 2011, 2014).

Otro hallazgo importante del estudio 2, fue confirmar que la edad no fue un factor determinante de las diferencias de eficacia del RI. Este resultado implica que, se generó nueva evidencia de que es posible estudiar diferencias de habilidad inferencial, a partir de contenidos STEM, desde los 4 años. Es decir, con base en la evidencia es posible afirmar que los niños participantes están en capacidad de expresar conclusiones precisas, coherentes y relevantes, además de hacer generalizaciones plausibles. Es probable que este conjunto de hallazgos represente un precedente importante para revisar el supuesto sobre el cambio característico en la trayectoria del desarrollo conceptual. Dicho supuesto se refiere, a que los niños mayores de 7 años hacen generalizaciones e inferencias basadas en conocimiento conceptual y los menores de 3, las hacen con base en características de la apariencia de las categorías, o sus propiedades superficiales, como plantearon Sloutsky y Fisher (2004).

Además, estos resultados aportan sustento a planteamientos como los de Hayes (2007), sobre la estabilidad en el desarrollo de los procesos cognitivos, fundamentales, involucrados en el RI entre los 4 y 6 años de edad. Desde esa perspectiva, los cambios más significativos en la inducción,

relacionados con la edad, pueden asociarse con la acumulación de conocimientos más precisos sobre diversas propiedades de los objetos y categorías. En consecuencia, las diferencias en los puntajes de las subescalas del IMERIN, por ejemplo, diferencias entre puntuaciones del dominio Tecnología y el de Ciencias, se pueden entender a partir de las particularidades de los conocimientos de los niños en cada área. Particularmente, con conocimiento de la existencia de propiedades invisibles, que tienen probabilidad de ser causales en algún sistema de relaciones.

Los hallazgos de la instrumentalización, resaltan la función de las tareas de razonamiento, multidimensionales, en la estimación de atributos de las diferencias psicológicas individuales, como lo explican Leighton y Sternberg, (2004); Sternberg y Gardner, (1983) y Vandekerckhove, (2014). Por ende, esta disertación aporta un método de medición válido y un conjunto de estrategias para la caracterización del RI temprano, confiables. Los cuales, pueden ser implementados para mejorar el aprendizaje de habilidades para razonar, eficientemente, a partir de diferentes tipos de datos, de conocimiento verificable y de evidencias sobre factores de cambio en los sistemas relacionales, del mundo real (Griffin et al., 2012; OECD, 2009, 2012a, 2013a; UNESCO, 2005, 2015, 2017). Se justifica la formación temprana de este tipo de habilidades, puesto que son precursoras de las competencias propias del pensamiento científico (Baillargeon et al., 2016; Hayes, 2007; Osborne, 2013; Ricco, 2015).

Alcances y limitaciones de los resultados de la investigación

Las adecuaciones de contenidos necesarias para continuar el proceso investigativo, a pesar de la crisis y el cierre de escuelas ocasionados por la Pandemia, condujeron a ampliar la documentación de experiencias de formación del pensamiento de los niños, en condiciones extraordinarias. Es probable que el resultado de dicha revisión, aporte nuevo conocimiento de estrategias para abordar problemas limitantes del acompañamiento docente. Incluso, podrían ser útiles para abordar soluciones viables, aun en ausencia de rutinas escolares para el aprendizaje conceptual. En ese sentido, los resultados de la actualización de referentes teóricos y pragmáticos, condujo al estudio de la influencia de prácticas pedagógicas familiares en el desarrollo de habilidades inferenciales de los participantes. Particularmente, a la indagación de la acción pedagógica intencional, ocurrida en ausencia del acompañamiento, o intervención de los agentes presentes en el aula de Preescolar.

Adicionalmente, en el proceso de reformulación de métodos y estudio de la literatura actualizada se descubrieron algunas limitaciones en la indagación de procesos cognitivos implicados en el aprendizaje de habilidades de inferencia probabilística. Este hallazgo, apoyó la sugerencia hecha a las docentes participantes, respecto a la necesidad de asumir una nueva perspectiva para la creación de indicadores de logro en el Preescolar. Por ejemplo, una que integre enfoques, tanto de la Pedagogía como de la Psicología, para intentar la transformación de intereses en formación del pensamiento científico temprano, como lo explica Samarapungava (1996; Gopnik, 2011).

Los resultados de la segunda etapa del estudio 1, fueron consecuencia de la actualización de marcos conceptuales y del ajuste a la metodología, dadas las restricciones a la escolaridad durante los años 2020 y 2021. Dichos resultados podrían aportar información y datos nuevos, para conciliar posiciones contrastantes respecto a la importancia de la acción pedagógica explícita, para el cambio conceptual durante los años de escolarización iniciales (Butler y Markman, 2012; Gelman, 2003; Lo et al., 2002; Kuhn, 2010; Nisbett, et al., 1982). Probablemente, este tipo de evidencia sea difícil de generar de nuevo, por lo cual este estudio sería de casos atípicos. A menos de que se repita el cierre de aulas, la indagación del RI temprano deberá realizarse con la participación de niños escolarizados en circunstancias de “normalidad”.

Los instrumentos para la indagación con las madres fueron producto del abordaje pragmático de los modelos informales de razonamiento. Dichos modelos plantean que los sistemas de pensamiento tienen origen en prácticas culturales diversas, por lo tanto, sugieren que los procesos de inferencia son sustancialmente diferentes entre los grupos humanos (Nisbett et al., 2001). De esa forma y a partir de los hallazgos de los estudios de la etapa 2 se verificó la hipótesis sobre la influencia de claves sociales, lingüísticas y pedagógicas en la eficacia de la inferencia inductiva en situaciones de razonamiento con contenidos de dominios STEM.

Dado que no se encontró literatura referente a estudios del desarrollo del RI eficaz con participantes mayores de 4 años (edad preescolar) desescolarizados, al menos en Colombia, se puede afirmar que esta investigación es pionera. Esta afirmación se sustenta en el análisis de la evidencia, aportada al cuerpo de conocimiento de las teorías de la socialización de la cognición, a partir del estudio 3 y sus resultados. Según lo expuesto por Holland y cols., (1989), esta clase de evidencias son necesarias para suplir falencias en los principios de las teorías sobre desarrollo cognitivo temprano. En ese orden de ideas, los resultados en sí mismos, son únicos respecto al

conocimiento de las particularidades de la influencia de la interacción pedagógica en el desarrollo conceptual, y a través de éste, en el aprendizaje de habilidades inferenciales de los niños participantes.

De otra parte, es posible que el instrumento de indagación de las prácticas de socialización de la cognición fuera diligenciado durante periodos de crisis emocionales intensas, ocasionadas por las problemáticas propias de la Pandemia. En consecuencia, es probable que los autoreportes de las madres fueran superficiales, espontáneos o descuidados, y no producto de la reflexión, ni de la consciencia sobre las dinámicas referidas en las preguntas. Esta posibilidad limitaría el alcance de la interpretación de resultados de la encuesta. Concretamente, en términos de suponer que existe evidencia, confiable, sobre los hábitos de búsqueda de conocimiento valorado en la Cultura.

Es importante mencionar que las limitaciones para la interpretación de resultados, derivadas de la crisis sanitaria durante 2020 y 2021, especialmente, a las posibilidades de proyección y de réplica de las indagaciones en este proceso investigativo, se deberán traducir en preguntas, hipótesis y nuevas cuestiones para estudiar. Concretamente, en relación con los factores culturales asociados a la búsqueda intencional del conocimiento STEM.

Conclusiones

Caracterizar el RI en la infancia temprana del siglo XXI es un desafío porque implica asumir una perspectiva del desarrollo cognitivo, integradora y multidisciplinaria. Esta perspectiva implica delimitar la función del RI en la transformación del conocimiento subjetivo en conocimiento verificable (probablemente objetivo). Además, tiene importancia crítica diseñar y seguir, sistemáticamente, una estrategia metodológica de indagación de la inferencia, en una amplia variedad de modalidades explícitas; para probar que el constructo tiene atributos mensurables. De hecho, caracterizar es el proceso metodológico, mínimo necesario, para sustentar que la capacidad de RI de los niños en edad preescolar se puede estudiar a partir de tareas de inferencia, cuyos contenidos pertenecen a dominios específicos (Goswami, 2010; Lakin y Gambrell 2012; Legare, 2014; Miller et al., 2012).

La medición realizada mediante el IMERIN permitió especificar cómo los niños integran su conocimiento conceptual, particularmente sobre causalidad en dominios específicos, en la comprensión de estructuras inductivas más generales. Lo anterior es un avance importante para los estudios de la trayectoria del desarrollo del RI. Este aporte se proyectará, usando las estrategias de

evaluación propuestas, para examinar si el proceso de cambio en las habilidades inferenciales se puede influir mediante la enseñanza-aprendizaje sobre probabilidad, covariación y azar, usando actividades de dominios STEM.

Un importante logro para la disertación fue establecer la importancia de la multidisciplinariedad, para la validez ecológica de las estrategias de evaluación e interpretación de resultados en la estimación de atributos de la inferencia inductiva. Desde lo metodológico, este logro dependió de: a) integrar a la creación de actividades de RI el aporte de científicos, Psicólogos, expertos en diferentes campos del conocimiento STEM y docentes profesionales en Educación Preescolar; y b) adaptar los objetivos de la evaluación de atributos de la inferencia, a las condiciones y modalidades que favorecen los desempeños de los niños más jóvenes, según lo reportan los antecedentes disponibles (Sobel et al., 2017).

Los resultados del estudio 1, contribuyeron a la consolidación de un conjunto de datos sobre funcionalidad del instrumento inicial, que fueron de importancia significativa para la creación de nuevas estrategias metodológicas para la medición requerida. Estos también resultaron confiables, a pesar de que no fue posible usar las actividades que habían sido creadas y probadas inicialmente. De hecho, los datos tienen el potencial de orientar futuras indagaciones del desarrollo del RI, dado que es altamente probable que representen evidencia faltante en los cuerpos de conocimiento de la Psicología Cognitiva, ya que se obtuvieron en el aula, con la participación de familias pertenecientes a estratos socioeconómicos bajos, y en las circunstancias de normalidad escolar previas a la Pandemia.

A pesar de las condiciones referidas, el supuesto que encaminó la indagación inicial se mantuvo en los estudios 2 y 3. Este orientó las pruebas de hipótesis que permitieron demostrar que, desde los 4 años, los niños están en capacidad de razonar eficazmente. Incluso, si las situaciones presentan condiciones de probabilidad en las relaciones o eventos razonados, dado que usan conocimiento y generalizaciones previos como fundamento de inferencias posteriores. Este principio es fundamental en la indagación del desarrollo cognitivo y ha sido demostrado empíricamente. Por ejemplo, en los trabajos revisados por Sutherland y Cimpian (2016), realizados con población de países desarrollados. Por lo tanto, al corroborar dicho supuesto, con participación de niños colombianos desescolarizados, se avanzó en la generación de conocimiento faltante entre los fundamentos de las teorías de desarrollo cognitivo, del siglo XXI (consolidadas a partir de estudios con poblaciones de países industrializados, escolarizadas).

Investigando la habilidad inferencial y creando instrumentos para su evaluación y diagnóstico desde el enfoque de la Educación STEM, fue posible proponer un índice de eficacia en el RI. El cual podría ser útil para la toma de decisiones respecto a la intervención psicológica, o psicopedagógica en Preescolar y para el diseño de proyectos culturales, educativos y de aula, *inclusivos*. Al respecto, es importante mencionar que esta proyección de resultados fue motivada por el interés en apoyar la sugerencia de varios autores de estudios del desarrollo cognitivo, ecológicamente válidos. Por ejemplo, Bogdan y Meneses (2019) sugieren indagar, con la participación de niños, el origen de la brecha de género en la elección de profesiones u ocupaciones en áreas STEM, y la necesidad de abordarla como un reto social a superar.

En ese contexto, puntajes de nivel básico en las dimensiones del IMERIN se pueden analizar como evidencia de la preferencia por algún tipo de representación semántica particular, o de sesgos en las creencias (críticamente influenciadas por la cultura). De hecho, es posible que, por asociación con el género del menor, se elija comunicar ciertas representaciones y tipos de evidencia (Butler y Markman, 2012), especialmente en el dominio Tecnología (UNESCO, 2017). Estas representaciones idealizadas se usan para razonar en todos los contextos, a partir de contenidos de diferentes dominios, como lo explica Foster (2020). No obstante, dicha preferencia puede influirse pedagógicamente, en los diferentes entornos de socialización cognitiva de los niños, incluida el aula de Preescolar.

Dicha evidencia también se halló en estudios como los de Lane y Shafto (2017); Samarapungavan et al., (2008); Setoh et al, 2013; y Shafto et al, (2008). Sin embargo, en los países en los cuales se realizaron las investigaciones referidas ya se había implementado el enfoque STEM en la Educación pública, desde el nivel Preescolar. En ese sentido, es posible suponer razonablemente, que los resultados del proceso de caracterización del RI representan un antecedente valioso que podría servir como referente en la construcción de propuestas de innovación pedagógica en Colombia y en otros países en los que aún no se reglamenta la Educación STEM.

El IMERIN tiene propiedades psicométricas adecuadas, de manera que cumple con criterios de validez y confiabilidad, definidos por la Teoría clásica de los Test. Por lo tanto, la medición que permite y la utilidad de sus resultados se ajustan a los estándares establecidos por las organizaciones AERA, APA y NCME, (2014) para las evaluaciones en Educación con fines diagnósticos. En ese orden de ideas, el proceso metodológico generó nueva evidencia empírica, de

la capacidad de los niños, en primera infancia, para razonar eficazmente sobre leyes y principios científicos. Validar la interpretación los resultados del IMERIN, específicamente, respecto a su utilidad científica, fue consecuencia de involucrar conceptos de los diferentes expertos, en la interpretación índices de desarrollo de habilidades inferenciales. La tendencia en dichos conceptos, fue proyectar los resultados hacia propuestas de renovación de estándares de formación del pensamiento infantil en la sociedad del conocimiento (Gopnik, 2012; Kuhn, 2010a; Ricco, 2015).

Respecto a la influencia de la mediación de las madres en el RI de los participantes, se descubrió que enseñan principios y reglas que regulan algunas relaciones causales. En especial, aquellas que siguen un patrón determinado. Sin embargo, no se reportaron experiencias de exploración de la causalidad probabilística, ni de condiciones estocásticas en sistemas de causalidad. Es probable que esta tendencia se pueda explicar a partir de ideas como las de Schulz y Sommerville (2006). Los autores plantearon, que las personas que forman el pensamiento de los niños podrían estar menos orientadas a considerar adecuada la flexibilidad cognitiva el razonamiento sobre causalidad, en contraste con el razonamiento sobre otros tipos de relaciones o covariaciones.

En ese sentido se recomienda, para los procesos de mediación en la búsqueda de conocimiento científico, integrar el conocimiento de las formadoras sobre causas no obvias y propiedades causales invisibles, a las explicaciones de la estructura física de eventos causales cotidianos, como las refieren Sobel y Legare (2014). Esto podría lograrse mediante experiencias de aprendizaje-enseñanza social, en cualquiera de los ámbitos de convivencia de los menores y haciendo uso de las TIC. Se esperaría, además, que las herramientas de medición e indagación fueran objeto de estudio en futuras investigaciones y en experiencias del aula y la comunidad del Preescolar. Quizás, desde el interés en contribuir en la formación del pensamiento científico de las nuevas generaciones.

Referencias

Aguiar, N. R., Stoess, C. J., y Taylor, M. (2012). The development of children's ability to fill the gaps in their knowledge by consulting experts. *Child Development*, 83, 1368- 1381. <https://doi.org/10.1111/j.1467-8624.2012.01782.x>

- Ahl, R. y Keil, F. (2017). Diverse Effects, Complex Causes: Children Use Information About Machines' Functional Diversity to Infer Internal Complexity. *Child Development*, 88 (3), 828-845. <https://doi.org/10.1111/cdev.12613>
- Anderson, R., y Branstetter, S. (2012). Adolescents, parents, and monitoring: A review of constructs with attention to process and theory. *Journal of Family Theory y Review*, 4, 1-19. <https://doi.org/10.1111/j.1756-2589.2011.00112.x>
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, United States: Author.
- Avila, C. y Barragan, G. (2018). Artículo de Experiencia en el Aula: Educación STEM una ruta hacia la innovación. *Revista electrónica TicALS*, 4, 146 - 162 www.als.edu.co/revistaticals
- Barkl, S., Porter, A., y Ginns, P. (2012). Cognitive training for children: Effects on inductive reasoning, deductive reasoning, and mathematics achievement in an Australian school setting. *Psychology In The Schools*, 49(9), 828-842. <https://doi.org/10.1002/pits.21638>
- Baillargeon, R., Scott, R. M., y Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67. doi.org/10.1146/annurev-psych-010213-115033
- Barbey, A. K., y Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30, 3, 241-254. <https://doi.org/10.1017/S0140525X07001653>
- Bascandzhev, I., y Harris, P.L. (2016). The beautiful and the accurate: Are children's selective trust decisions biased?. *Journal of Experimental Child Psychology*, Vol. 152, 92-105, doi.org/10.1016/j.jecp.2016.06.017
- Batanero, C., y Chernoff, E. J. (2018). *Teaching and Learning Stochastics: Advances in Probability Education Research*. <https://doi.org/10.1007/978-3-319-72871-1>
- Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., y Sánchez, E. (2016). *Research on Teaching and Learning Probability*. Cham: Springer International Publishing.
- Bogdan Toma, R. y Meneses Villagrà, J. Á. (2019). Preferencia por contenidos científicos de física o de biología en Educación Primaria: un análisis clúster. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, 16(1),1, http://unal.edu.co/10.25267/Rev_Eureka_ensen_divulg_cienc.2019.v16.i1.1104

- Bolt, D. (2007). The Present and Future of IRT-Based Cognitive Diagnostic Models (ICDMs) and Related Methods. *Journal Of Educational Measurement*, (4). 377.
- Bonawitz, E. B., y Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, 48, 4, 1156-1164. <https://doi.org/10.1037/a0026471>
- Bonawitz, E., Ullman, T. D., Bridgers, S., Gopnik, A., y Tenenbaum, J. B. (2019). Sticking to the Evidence? A Behavioral and Computational Case Study of Micro-Theory Change in the Domain of Magnetism. *Cognitive Science*, 43(8). <https://doi.org/10.1111/cogs.12765>
- Bonett, D.G., Wright, T.A. Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika* 65, 23–28 (2000). <https://doi.org/10.1007/BF02294183>
- Bonnefon J-F y Billaut E. (2016). Individual Differences in Reasoning beyond Ability and Disposition Cap. 11 En Macchi, Laura, Bagassi, Maria y Viale, Riccardo. *Cognitive unconscious and human rationality*. Toppan Best-set Premedia Limited. United States of America. Massachusetts Institute of Technology
- Bouwmeester, S., y Sijtsma, K. (2004). Measuring the ability of transitive reasoning, using product and strategy information. *Psychometrika*, 69(1), 123. <https://doi.org/10.1007/BF02295843>
- Borovenik M., Kapadia R. (2018) Reasoning with Risk: Teaching Probability and Risk as Twin Concepts. En: Batanero C., Chernoff E. (eds) *Teaching and Learning Stochastics*. ICME-13 Monographs. Springer, Cham.
- Badger, J. R., y Shapiro, L. R. (2012). Evidence of a transition from perceptual to category induction in 3- to 9-year-old children. *Journal of Experimental Child Psychology*, 113(1), 131–146. <https://doi.org/10.1016/j.jecp.2012.03.004>
- Brandone, A. C. (2017). Changes in Beliefs About Category Homogeneity and Variability Across Childhood. *Child Development*, 88(3), 846-866. <https://doi.org/10.1111/cdev.12616>
- Bramley, N., Gerstenberg, T., Tenenbaum, J. y Gureckis, T. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, 105, 9-38. <https://doi.org/10.1016/j.cogpsych.2018.05.001>
- Brenneman, K., y Louro, I. F. (2008). Science Journals in the Preschool Classroom. *Early Childhood Education Journal*, 36(2), 113–119. [https://doi.org/10.1007/s10643-008-0258-](https://doi.org/10.1007/s10643-008-0258-z)

- Bryant, P. y Nunes, T. (2012). *Children's understanding of probability: A literature review*. London, England: Nuffield Foundation.how
- Bulloch, M. J. y Opfer, J. E. (2009). What makes relational reasoning smart? Revisiting the perceptual-to-relational shift in the development of generalization. *Developmental Science*, 12, 1, 114-122. <https://doi.org/10.1111/j.1467-7687.2008.00738.x>
- Bumeltshammer, K.S., y Kirkham, N.Z. (2013). Learning to look: probabilistic variation and noise guide infants' eye movements. *Developmental Science*, 16, 760-771. <https://doi.org/10.1111/desc.12064>
- Burris, V. (1982). The Child's Conception of Economic Relations; A Study of Cognitive Socialization. *Sociological Focus*, 15(4), 307-325. <https://doi.org/10.1080/00380237.1982.10570424>
- Butler, L. P., y Markman, E. M. (2012). Preschoolers Use Intentional and Pedagogical Cues to Guide Inductive Inferences and Exploration. *Child Development*, 83(4), 1416-1428. <https://doi.org/10.1111/j.1467-8624.2012.01775.x>
- Cánavos, G. C. (1999). *Probabilidad y Estadística: Aplicaciones y Métodos*. México: McGraw-Hill.
- Carey, S. (2009). *The Origin of Concepts*. Oxford Series in Cognitive Development.
- Clark, E. V. (2004). *How language acquisition builds on cognitive development*. *Trends in Cognitive Sciences*, 8(10), 472-478. <https://doi.org/10.1016/j.tics.2004.08.012>
- Cerchiaro-Ceballos, E., y Puche-Navarro, R. (2018). Funcionamientos inferenciales en niños caminadores: un acercamiento al microdesarrollo en una tarea de resolución de problemas. *Revista Colombiana de Psicología*, 27, 117-135. <https://doi.org/10.15446/rcp.v27n2.66054>
- Csapo, B. (1997). The Development of Inductive Reasoning: Cross-sectional Assessments in an Educational Context. *International journal of behavioral development*, (4). 609. <https://journals.sagepub.com/doi/10.1080/016502597385081>
- Cosmides L, Tooby J. (2013). Evolutionary psychology: New perspectives on cognition and motivation. *Annu. Rev. Psychol.* 64:201-29. DOI: 10.1146/annurev.psych.121208.131628
- Colberg, M., Nester, M. A., y Cormier, S. M. (1982). Inductive reasoning in psychometrics: A philosophical corrective. *Intelligence*, 6139-164. [https://doi.org/10.1016/0160-2896\(82\)90011-3](https://doi.org/10.1016/0160-2896(82)90011-3)

- Colberg, M., Nester, M., y Trattner, M. (1985). Convergence of the Inductive and Deductive Models in the Measurement of Reasoning Abilities. *Journal Of Applied Psychology*, 70(4), 681-694. <https://doi.org/10.1037/0021-9010.70.4.681>
- Corral, Y. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista Ciencias de la Educación*, 19(33), 228-247. <https://bit.ly/1T1z0ct>
- Culbertson M. (2016). Bayesian Networks in Educational Assessment: The State of the Field. *Applied Psychological Measurement*, 40(1) 3–21 <https://doi.org/10.1177/0146621615590401>
- Cummins D. (2004). The Evolution of Reasoning. En Leighton, J.P. y Sternberg R. J. (Ed.).(2004). *The nature of reasoning*. Cambridge University Press.
- Damon, W. (1990). Social Relations and Childrens Thinking Skills. Contributions to Human Development. En Kuhn D (ed). *Developmental Perspectives on Teaching and Learning Thinking Skills*, pp 95–107. <https://doi.org/10.1159/000418983>
- Darling, N., y Steinberg, L. (1993). *Parenting style as context: An integrative model*. *Psychological Bulletin*, 113(3), 487–496. <https://doi.org/10.1037/0033-2909.113.3.487>
- Dasgupta, I., Schulz, N., Goodman, D. y Gershman, S. (2018). Remembrance of inferences past: Amortization in human hypothesis generation, *Cognition*, 178, 67-81. <https://doi.org/10.1016/j.cognition.2018.04.017>
- De Koning, E., Sijtsma, K., y Hamers, J. M. (2003). Construction and Validation of a Test for Inductive Reasoning. *European Journal Of Psychological Assessment*, (1). 24. <https://doi.org/10.1027//1015-5759.19.1.24>
- Denison, S. y Xu, F. (2012). Probabilistic Inference in Human Infants. En T. Kushnir, y F. Xu (Eds.), *Advances in child development and behavior: Rational constructivism in cognitive development*. Academic Press, Elsevier. doi.org/10.1016/B978-0-12-397919-3.00002-2
- Dean Jr., D., y Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, 91(3), 384–397. <https://doi.org/10.1002/sce.20194>
- Dienes, Z., y Perner, J. (1999) A theory of implicit and explicit knowledge. *Behavioural and Brain Sciences*, 22,735-755. DOI: 10.1017/s0140525x99002186
- Denison, S., y Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, 130, 335–347. <https://doi.org/10.1016/j.cognition.2013.12.001>

- Duque Aristizábal, C. P., Aristizábal, C. P. D., Márquez, Á. V. V., y Gutiérrez, A. P. H. (2010). Comprensión inferencial de textos narrativos en primeros lectores: una revisión de la literatura. *Ocnos: Revista De Estudios Sobre Lectura*, (6), 35. https://doi.org/10.18239/OCNOS_2010.06.03
- Ebersbach, M., y Resing, W. M. (2008). Implicit and Explicit Knowledge of Linear and Exponential Growth in 5- and 9-Year-Olds. *Journal Of Cognition y Development*, 9(3), 286-309. <https://doi.org/10.1080/15248370802247962>
- Einav, S. y Robinson E. J. (2010). Children's sensitivity to error magnitude when evaluating informants. *Cognitive Development* 25, 218–232, <https://doi.org/10.1016/j.cogdev.2010.04.002>
- Eichler, A., y Vogel, M. (2012). Basic modelling of uncertainty: Young students' mental models. *ZDM Mathematics Education*, 44(7), 841–854. <https://doi.org/10.1007/s11858-012-0451-9>.
- Elosua Oviden, P. y Zumbo, B.D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901 <http://www.redalyc.org/articulo.oa?id=72720458>
- Epstein, N., y Fischer, M. R. (2017). Academic career intentions in the life sciences: Can research self-efficacy beliefs explain low numbers of aspiring physician and female scientists?. *PLOS ONE*, 12(9), e0184543. <https://doi.org/10.1371/journal.pone.0184543>
- Erickson, J. E., Keil, F. C., y Lockhart, K. L. (2010). Sensing the Coherence of Biology in Contrast to Psychology: Young Children's Use of Causal Relations to Distinguish Two Foundational Domains. *Child Development*, 81(1), 390-409. <https://doi.org/10.1111/j.1467-8624.2009.01402.x>
- Evans, J. St. B. T. (2020). *Hypothetical thinking: Dual processes in reasoning and judgment*. Psychology Press and Routledge Classic Editions.
- Evans, J. St. B. T., y Over, D. E. (2013). Reasoning to and from belief: Deduction and induction are still distinct. *Thinking & Reasoning*, 19 (3), 267-283. <https://cogentoa.tandfonline.com/doi/full/10.1080/13546783.2012.745450>
- Ferrar, S. J., Stack, D. M., Dickson, D. J., Serbin, L. A., Ledingham, J., y Schwartzman, A. E. (2019). Maternal Socialization Responses to Preschoolers' Success and Struggle: Links to

- Contextual Factors and Academic and Cognitive Outcomes. *Journal of Research in Childhood Education*, 1–19. <https://doi.org/10.1080/02568543.2019.1607787>
- Fiorini, M., y Keane, M. P. (2014). How the Allocation of Children’s Time Affects Cognitive and Noncognitive Development. *Journal of Labor Economics*, 32(4), 787–836. <https://doi.org/10.1086/677232>
- Fisher, A. V., Godwin, K. E., Matlen, B. J. y Unger, L. (2015). Development of category-based induction and semantic knowledge. *Child Development*, 86(1), 48–62. <https://doi.org/10.1111/cdev.12277>
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 5, 28–45. <http://dx.doi.org/10.14786/flr.v2i3.96>
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht; Reidel Publishing Company.
- Fizke, E., Butterfill, S., Van de Loo, L., Reindl, E., y Rakoczy, H. (2017). Are there signature limits in early theory of mind?. *Journal Of Experimental Child Psychology*, 209. <https://doi.org/10.1016/j.jecp.2017.05.005>
- Foster-Hanson E, Moty K, Cardarelli A, Ocampo JD, y Rhodes M. (2020). Developmental Changes in Strategies for Gathering Evidence About Biological Kinds. *Cogn Sci*. 2020 May; 44 (5). <https://doi.org/10.1111/cogs.12837>.
- Frost, R., Armstrong, B. C., y Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, 145(12), 1128–1153. doi.org/10.1037/bul0000210
- Gandhi H. (2018). Understanding Children’s Meanings of Randomness in Relation to Random Generators. En: Batanero, C., y Chernoff, E. J. (Eds.). (2018). *Teaching and Learning Stochastics: Advances in Probability Education Research*. doi.org/10.1007/978-3-319-72871-1
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, 20, 65–96. [https://doi.org/10.1016/0010-0285\(88\)90025-4](https://doi.org/10.1016/0010-0285(88)90025-4)
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday life*. New York, NY: Oxford University Press.

- Gelman, R., y Brenneman, K. (2012). Moving young “scientists-in-waiting” onto science learning pathways: Focus on observation. In J. Shrager y S. Carver (Eds.), *The journey from child to scientist: Integrating cognitive development and the education sciences*. 155–169. American Psychological Association. <https://doi.org/10.1037/13617-008>
- Gelman, S. A., y Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, (3), 327. <https://doi.org/10.1016/j.cogpsych.2013.02.001>
- Gelman, S. A., Leslie, S.-J., Was, A. M., y Koch, C. M. (2015). Children’s interpretations of general quantifiers, specific quantifiers and generics. *Language, Cognition, and Neuroscience*, 30, 448–461. <https://doi.org/10.1080/23273798.2014.931591>
- Gennari, S., Sloman S., Malt, B., y Fitch, W. (2002). Motion events in language and cognition. *Cognition*, 83 (1), 49 – 79. DOI:10.1016/S0010-0277(01)00166-4
- Gigerenzer, G. (2015). *Calculated risks: How to know when numbers deceive you*. New York: Simon y Schuster Audio.
- Gil Chaves, L., y Flórez Romero, R. (2013). Desarrollo de habilidades de pensamiento inferencial y comprensión de lectura en niños de tres a seis años. *Panorama*, 5(9). <https://doi.org/10.15765/pnrm.v5i9.39>
- Giroto, V., y Gonzalez, M. (2007). How to elicit sound probabilistic reasoning: Beyond word problems. *Behavioral and Brain Sciences*, 30, 3, 268. doi 10.1017_S0140525X07001768
- Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Phil. Trans. R. Soc. B*, 372 (1711). doi.org/10.1098/rstb.2016.0054.
- Gopnik, A. (2011). The Theory of Theory 2.0: Probabilistic Models and Cognitive Development. *Child Development Perspectives*. 5, 3, 161-163, [HTTPS://DOI.ORG/10.1111/j.1750-8606.2011.00179.x](https://doi.org/10.1111/j.1750-8606.2011.00179.x)
- Gopnik, A. (2012). Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications. *Science*, Sep. 28;337(6102): 1623-7. <https://doi.org/10.1126/science.1223416>
- Gopnik, A., y Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, 10(3), 281–287. <https://doi.org/10.1111/j.1467-7687.2007.00584.x>

- Gopnik, A., Sobel, D.M., Schulz L.E., Glymour C. (2001). Causal Learning Mechanisms in Very Young Children: Two-, Three-, and Four-Year-Olds Infer Causal Relations From Patterns of Variation and Covariation *Developmental Psychology* . 37, (5), 620-629. <https://doi.org/10.1037//0012-1649.37.5.620>
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., y Danks, D. (2004). *A Theory of Causal Learning in Children: Causal Maps and Bayes Nets*. *Psychological Review*, 111(1), 3–32. <https://doi.org/10.1037/0033-295x.111.1.3>
- Goswami, U. (ed.). (2003). Wiley-Blackwell handbook of childhood cognitive development. New York: Wiley
- Goswami, U. (Ed.). (2010). Wiley-Blackwell handbook of childhood cognitive development. (2nd ed.). Wiley.
- Granovskiy, B. (2018). Science, Technology, Engineering, and Mathematics (STEM) Education: An Overview. Code R45223; Congressional Research Service, USA. www.crs.gov
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A.C. y Martin, R. (2014). Domain-general problem solving skills and education in the 21st century, *Educational Research Review* <https://doi.org/http://dx.doi.org/10.1016/j.edurev.2014.10.002>
- Griffiths, T. L., Tenenbaum, J. B., y Kemp, C. (2012). Bayesian inference. *The Oxford handbook of thinking and reasoning*, 22-35.
- Griffin, P., McGaw, B., y Care, E. (Eds.) (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer.
- Grigorenko E., Sternberg R.J., y Strauss, S. (2006). Practical intelligence and elementary-school teacher Thinking Skills and Creativity Volume 1 issue 1 doi 10.1016_j.tsc.2005.03.001
- Gweon, H., Tenenbaum, J. B., y Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107(20), 9066–9071. www.pnas.org/content/pnas/107/20/9066.full.pdf
- Hasson U. (2017). The neurobiology of uncertainty: implications for statistical learning. *Phil. Trans. R. Soc. B*, 372 (1711). [https://doi.org/doi 10.1098_rstb.2016.0048](https://doi.org/doi%2010.1098_rstb.2016.0048)
- Hayes, B. K., McKinnon, R., y Sweller, N. (2008). *The development of category-based induction: Reexamining conclusions from the induction then recognition (ITR) paradigm*.

- Developmental Psychology*, 44(5), 1430–1441. <https://doi.org/10.1037/0012-1649.44.5.1430>
- Hayes, B. K., Stephens, R. G., Ngo, J., y Dunn, J. C. (2018). The Dimensionality of Reasoning: Inductive and Deductive Inference can be Explained by a Single Process. *Journal Of Experimental Psychology. Learning, Memory, And Cognition*, <https://doi.org/10.1037/xlm0000527>
- Heller, J., Stefanutti, L., Anselmi, P., y Robusto, E. (2015). On the Link between Cognitive Diagnostic Models and Knowledge Space Theory. *Psychometrika*, (4), 995. <https://doi.org/10.1007/s11336-015-9457-x>
- Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin y Review*, 7, 569– 592. <https://doi.org/10.3758/bf03212996>
- Hernández I. (2010). Factores psicosociales asociados al alto nivel de logro académico de dos Estudiantes sensibles al medio. Tesis presentada para optar al Título de Magíster. Repositorio institucional de la Universidad Pedagógica Nacional. Colombia.
- Hernández I. y Pacheco P. (2013). Experiencias de innovación didáctica en la enseñanza de la Estadística. En: Inclusión social con responsabilidad, reformas educativas y profesionalización del Docente- Tomo III de la colección “Investigación Educativa y Políticas Públicas”. Universidad de Nayarit- México. ISBN 978-607-7868-70-5. 2014.
- Herro, D., y Quigley, C. (2016). Exploring teachers’ perceptions of STEAM teaching through professional development: implications for teacher educators. *Professional Development in Education*, 43(3), 416–438. <https://doi.org/10.1080/19415257.2016.1205507>
- Holland, J., Holyoak, K., Nisbett, R. y Thagard, P. (1989). *Induction Processes of Inference, Learning, and Discovery*. Michigan; The MIT Press.
- Howe, C. Nunes, T. y Bryant, P. (2011). Rational Number and Proportional Reasoning. *International Journal of Science and Mathematics Education*, 9, 2, 391-417. doi 10.1007_s10763-010-9249-9pil
- Ifenthaler, D., y Seel, N. M. (2011). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction*, 21(4), 538–549. <https://doi.org/10.1016/J.LEARNINSTRUC.2010.08.004>

- Jacobs, J. E., y Narloch, R. H. (2001). Children's use of sample size and variability to make social inferences. *Journal of Applied Developmental Psychology*, 22(3), 311–331. [https://doi.org/10.1016/s0193-3973\(01\)00086-7](https://doi.org/10.1016/s0193-3973(01)00086-7)
- Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., y Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, 14. <https://doi.org/10.1016/j.cognition.2015.03.00>
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 71-115. <https://www.sciencedirect.com/science/article/pii/S0364021381800055>
- Johnson, S. G. B., y Ahn, W. (2017). Causal mechanisms. En M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning*, p 127-146. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.013.12>
- Kahneman, D., Slovic, P. y Tversky, A. (eds.) (1982). *Judgment under Uncertainty, Heuristics and Biases*. Cambridge University Press.
- Klauer, K. J., Willmes, K., y Phye, G. D. (2002). Inducing Inductive Reasoning: Does It Transfer to Fluid Intelligence?. *Contemporary Educational Psychology*, (1). 1. <https://doi.org/10.1006/ceps.2001.1079>
- Keith, T. Z., y Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology In The Schools*, 47(7), 635-650. DOI:10.1002/pits.20496
- Kelly, A., Lesh, A. y Baek J. (eds) (2010). *Handbook of Design Research Methods in Education_ Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*. Mahwah, New York
- Kemp, C., y Jern, A. (2014). A taxonomy of inductive problems. *Psychonomic Bulletin y Review*, 21(1), 23-46. <https://doi.org/10.3758/s13423-013-0467-3>
- Kim, S., Kalish, C., Harris, P. (2012). Speaker reliability guides children's inductive inferences about novel properties, *Cognitive Development*, 27, (2), 114-125, <https://doi.org/10.1016/j.cogdev.2011.10.004>.
- Kinnear, V., y Clark, J. (2014). Probabilistic reasoning and prediction with young children. In J. Anderson, M. Cavanagh, y A. Prescott (Eds.), *Curriculum in focus: Research guided practice* (pp. 335–342). MERGA.

- Koenig, M. A., Cole, C. A., Meyer, M., Ridge, K. E., Kushnir, T., y Gelman, S. A. (2015). *Reasoning about knowledge: Children's evaluations of generality and verifiability. Cognitive Psychology*, 83, 22–39. <https://doi.org/10.1016/j.cogpsych.2015.08.007>
- Köksal-Tuncer, Ö., y Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, 48, 135–145. [doi-org.ezproxy.unal.edu.co/10.1016/j.cogdev.2018.06.011](https://doi.org/10.1016/j.cogdev.2018.06.011)
- Koslowski B. y Masnick A. (2010). Causal Reasoning and Explanation. En: U. Goswami, (Series Ed.). *Wiley-Blackwell handbook of childhood cognitive development*, (2nd ed.), pp. 341 – 419. New York : Wiley.xu
- Kühhirt, M. y Klein, M. (2018). Early Maternal Employment and Children's Vocabulary and Inductive Reasoning Ability: A Dynamic Approach. *Child Development*, 89 (2), 91-106. <https://doi.org/10.1111/cdev.12796>
- Kushnir, T., Wellman, H. M., y Gelman, S. A. (2008). *The role of preschoolers' social understanding in evaluating the informativeness of causal interventions. Cognition*, 107(3), 1084–1092. <https://doi.org/10.1016/j.cognition.2007.10.004>
- Kushnir, T., Wellman, H. M., y Gelman, S. A. (2009). *A self-agency bias in preschoolers' causal inferences. Developmental Psychology*, 45(2), 597–603. <https://doi.org/10.1037/a0014727>
- Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, 94(5), 810–824. <https://doi.org/10.1002/sce.20395>
- Kuhn D. (2010). What is Scientific Thinking and How Does it Develop?. En: U. Goswami, (Series Ed.). *Wiley-Blackwell handbook of childhood cognitive development*, (2nd ed., pg. 341 – 419). New York: Wiley.
- Kuhn, D., y Dean, D. J. (2004). Connecting Scientific Reasoning and Causal Inference. *Journal of Cognition and Development*, 5, 2, 261-288. https://doi.org/10.1207/s15327647jcd0502_5
- Kuhn, D. y Franklin S. (2006). The Second Decade: What Develops (And How), En: W. Damon (Editor-in-Chief), D. Kuhn y R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pg. 953–993). New York: Wiley.
- Kuhn, D., Iordanou, K., Pease, M. y Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking?. *Cognitive Development*, 23, 435–451. [doi-org.ezproxy.unal.edu.co/10.1016/j.cogdev.2008.09.006](https://doi.org/10.1016/j.cogdev.2008.09.006)

- Kuhn, D., Ramsey, S., y Arvidsson, T. S. (2015). Developing multivariable thinkers. *Cognitive Development*, 35, 92-110. doi 10.1016_j.cogdev.2014.11.003
- Kuhn, D., Schauble, L., y Garcia-Mila, M. (1992). *Cross-Domain Development of Scientific Reasoning. Cognition and Instruction*, 9(4), 285–327. https://doi.org/10.1207/s1532690xci0904_1
- Kushnir, T., Xu, F., y Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, 21, 1134–1140.
- Lakin, J. y Gambrell, J. (2012). Distinguishing verbal, quantitative, and figural facets of fluid intelligence in young students, *Intelligence*, 40, 6, 560-570, doi.org/10.1016/j.intell.2012.07.005
- Lane, J. D., Ronfard, S., Francioli, S. P., y Harris, P. L. (2016). Children’s imagination and belief: Prone to flights of fancy or grounded in reality? *Cognition*, 152, 127–140. <http://dx.doi.org/10.1016/j.cognition.2016.03.022>
- Lane, J. y Shafto, P. (2017). Young children’s attributions of causal power to novel invisible entities. *Journal of Experimental Child Psychology*, 162, 268-281, <doi.org/10.1016/j.jecp.2017.05.015>.
- Lawson, C. A. (2014). Three-year-olds obey the sample size principle of induction: The influence of evidence presentation and sample size disparity on young children’s generalizations. *Journal of Experimental Child Psychology*, 123, 147–154. <https://doi.org/10.1016/j.jecp.2013.12.004>
- Lawson, C. (2017). The Influence of Task Dynamics on Inductive Generalizations: How wSequential and Simultaneous Presentation of Evidence Impacts the Strength and Scope of Property Projections. *Journal Of Cognition y Development*, 18 (4), 493-513. <https://doi.org/10.1080/15248372.2017.1339707>
- Legare, C.H. (2014). The Contributions of Explanation and Exploration to Children's Scientific Reasoning. *Child Development Perspectives*. Vol. 8, (2) 101-106. <HTTPS://DOI.ORG/10.1111/cdep.12070>
- Leighton, J.P. y Sternberg R. J. (Ed.).(2004). *The nature of reasoning*. Cambridge University Press.
- Leighton, J. P., y Gierl, M. J. (2007). Defining and Evaluating Models of Cognition Used in Educational Measurement to Make Inferences about Examinees' Thinking Processes. *Educational Measurement: Issues And Practice*, 26(2), 3-16.

- Legare, C. H., y Lombrozo, T. (2014). *Selective effects of explanation on learning during early childhood*. *Journal of Experimental Child Psychology*, 126, 198–212. <https://doi.org/10.1016/j.jecp.2014.03.001>
- Lo, Y., Sides, A., Rozelle, J. y Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgments. *Cognitive Science - A Multidisciplinary Journal*, 26, 181-206. www.elsevier.com/locate/cogsci
- Lombrozo, T. (2016). Explanatory Preferences Shape Learning and Inference. *Trends in Cognitive Sciences*, 20(10), 748–759. <https://doi.org/10.1016/j.tics.2016.08.001>
- Low, J. y Perner J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, (2012), 30, 1–13. [HTTPS://DOI.ORG/10.1111/j.2044-835X.2011.02074.x](https://doi.org/10.1111/j.2044-835X.2011.02074.x)
- Lundberg, S. (2020). Educational gender gaps. *Southern Economic Journal*, Volume: 87 Issues: 2 416—439. [HTTPS://DOI.ORG/ 10.1002/soej.12460](https://doi.org/10.1002/soej.12460)
- Markovits, H., y Barrouillet, P. (2004). Introduction: Why is understanding the development of reasoning important?. En *Thinking y Reasoning*, 10(2), 113–121. <https://doi.org/10.1080/13546780442000006>
- Marcovitch, S. y Lewkowicz, D. J. (2009), Sequence learning in infancy: the independent contributions of conditional probability and pair frequency information. *Developmental Science*, 12: 1020-1025. <https://doi.org/10.1111/j.1467-7687.2009.00838.x>
- McGuinness Institute, Wellington, New Zealand. (2015). The future of scientific thought, *Journal of the Royal Society of New Zealand*, 45 (2), 95-100, <https://doi.org/10.1080/03036758.2015.1013142>
- McCormack, T., Bramley, N., Frosch, C., Fiona, P. y Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, 141, 1–22. [dx.doi.org/10.1016/j.jecp.2015.06.017](https://doi.org/10.1016/j.jecp.2015.06.017)
- Medin, D. L., Coley, J. D., Storms, G., y Hayes, B. L. (2003). *A relevance theory of induction*. *Psychonomic Bulletin y Review*, 10(3), 517–532. <https://doi.org/10.3758/bf03196515>
- Mercer, N. (2013). The Social Brain, Language, and Goal-Directed Collective Thinking: A Social Conception of Cognition and Its Implications for Understanding How We Think, Teach, and Learn. *Educational Psychologist*, 48(3), 148–168. <https://doi.org/10.1080/00461520.2013.804394>

- Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., y Kerns, K. A. (2012). A Latent Variable Approach to Determining the Structure of Executive Function in Preschool Children. *Journal Of Cognition y Development*, 13(3), 395-423. <https://doi.org/10.1080/15248372.2011.585478>
- Ministerio de Educación Nacional. (2009). *Desarrollo infantil y competencias en la primera infancia. Colombia*. Documento N° 10, MEN.
- Ministerio de Educación Nacional. (2011). *Valoración del desarrollo en contextos educativos*. Documento base, MEN: Colombia.
- Ministerio de Educación Nacional (2012). *Una propuesta de trabajo pedagógico en la educación inicial, documento base para la construcción del lineamiento pedagógico de educación inicial*. Documento de trabajo, MEN: Colombia.
- Mohlhenrich, E., Samsonau, S. y Spencer, R. (2018). *Integrating Science Through Authentic Research in Secondary Schools. IEEE Integrated STEM Education Conference (ISEC) - Princeton, NJ, USA* . doi 10.1109%2FISECon.2018.8340465
- Moore, B. (2009). Emotional Intelligence for School Administrators: A Priority for School Reform? *American Secondary Education*, 37(3), 20–28. <http://www.jstor.org/stable/41406313>
- Moore, J.D., Dorofy, P., Jabot, M., y Bouaynaya N. (2018). *Earth SySTEM: Investigating Earth from Space, IEEE Integrated STEM Education Conference (ISEC) - Princeton, NJ, USA*. doi 10.1109%2FISECon.2018.8340492
- Morewedge, C.K. y Kahneman D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, Vol. 14 (10), 435-440
- Moshman, D. (2004). From inference to reasoning: The construction of rationality. *Thinking y Reasoning*, 10(2), 221–239. <https://doi.org/10.1080/13546780442000024>
- Moulines, U. (2011). *El desarrollo moderno de la Filosofía de la Ciencia (1890-2000)*. México: Instituto de Investigaciones Filosóficas UNAM.
- Muniz, M., Seabra, A. G., y Primi, R. (2012). Validity and Reliability of the Inductive Reasoning Test for Children - IRTC. *Psicologia-Reflexao E Critica*, 25(2), 275-285.
- Nancekivell, S.H. y Friedman, O. (2017). She Bought the Unicorn From the Pet Store: Six- to Seven-Year-Olds Are Strongly Inclined to Generate Natural Explanations. *Dev. Psychol.* Vol. 53, No. 6, <https://doi.org/1079-1087/dev0000311>

- National Science Foundation. (2007). Women, Minorities, and Persons with Disabilities in Science and Engineering. National Science Foundation, Division of Science Resource Statistics. Arlington: U.S.A.
- Nguyen, S. P. (2012). Inductive Selectivity in Children's Cross-Classified Concepts. *Child Development*, 83(5), 1748-1761. <https://doi.org/10.1111/j.1467-8624.2012.01812.x>
- Nguyen, S., Gordbron C., Chevalier, T. y Girgis, H. (2016). Trust and doubt: An examination of children's decision to believe what they are told about food, *Journal of Experimental Child Psychology*, 144, 66-83. [dx.doi.org/10.1016/j.jecp.2015.10.015](https://doi.org/10.1016/j.jecp.2015.10.015)
- Nisbett, R. E., Krantz, D. H., Jepson, C., y Kunda, Z. (1983). *The use of statistical heuristics in everyday inductive reasoning*. *Psychological Review*, 90(4), 339–363. <https://doi.org/10.1037/0033-295x.90.4.339>
- Nisbett, R. y Ross, L. (1986). *Human Inference: Strategies And Shortcomings Of Social Judgment*. New Jersey: Prentice-Hall.
- Nisbett, R. E., Choi, I., Peng, K., y Norenzayan, A. (2001). Culture and Systems of Thought: Holistic Versus Analytic Cognition. *Psychological Review*, (2), 291-310. <https://doi.org/10.1037//0033-295X.108.2.291>
- Nisbett, R., Krantz, D., Jepson, C. y Fong G. T. (1982). Improving inductive inference. En Kahneman, D., Paul Slovic, Amos Tversky (eds.) (1982). *Judgment under Uncertainty, Heuristics and Biases*. Cambridge University Press.
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D., et al. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67, 2, 130–159 [doi 10.1037/a0026699](https://doi.org/10.1037/a0026699).
- Noyes, A., y Christie, S. (2016). Children Prefer Diverse Samples for Inductive Reasoning in the Social Domain. *Child Development*, 87(4), 1090-1098. <https://doi.org/10.1111/cdev.12522>
- OECD. (2009). *Programme for International Student Assessment. An overview*. Paris: OECD.
- OECD (2012a). *Better skills, better jobs, better lives. A strategic approach to skills policies*. Paris: OECD.
- OECD (2013a). *PISA 2012 assessment and analytical framework*. Paris: OECD Publishing.
- OECD. (2015) *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. PISA: OECD Publishing. <https://doi.org/10.1787/9789264229945-en>

- Opfer, J. E., y Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition*, 105, 206–217. <https://doi.org/10.1016/j.cognition.2006.08.006>
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265-279. [dx.doi.org/10.1016/j.tsc.2013.07.006](https://doi.org/10.1016/j.tsc.2013.07.006)
- Palejwala, M. H., y Fine, J. G. (2015). Gender differences in latent cognitive abilities in children aged 2 to 7. *Intelligence*, 4896-108. <https://doi.org/10.1016/j.intell.2014.11.004>
- Palmquist, C. M., y Jaswal, V. K. (2015). Preschoolers' inferences about pointers and labelers: The modality matters. *Cognitive Development*, 35, 178–185. <https://doi.org/10.1016/j.cogdev.2015.06.003>
- Pavlidou, E. V., y Bogaerts, L. (2019). Implicit Statistical Learning across Modalities and its Relationship with Reading in Childhood. *Front Psychol.* <https://doi.org/10.3389/fpsyg.2019.01834>
- Perner, J. y Dienes, Z. (2002). Implicit Versus Explicit Representation and Intra- Versus Inter-Modular Processing. *Computational Intelligence*, 18: 55-58. <https://doi.org/10.1111/1467-8640.00182>
- Perret, P. (2015). Children's Inductive Reasoning: Developmental and Educational perspectives. *Journal of Cognitive Education and Psychology*, 14 (3), 389 - 408. 10.1891/1945-8959.14.3.389
- Primi, R., Rocha da Silva, M. C., Rodrigues, P., Muniz, M., y Almeida, L. S. (2013). The use of the bi-factor model to test the uni-dimensionality of a battery of reasoning tests. *Psicothema*, 25(1), 115-122. <https://doi.org/10.7334/psicothema2011.393>
- Pavlidou, E., y Bogaerts, L. (2019). Implicit statistical learning across modalities and its relationship with reading in childhood. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2019.01834>
- Perret, P. (2015). Children's Inductive Reasoning: Developmental and Educational perspectives. *Journal of Cognitive Education and Psychology*, 14 (3), 389 - 408. <https://doi.org/10.1891/1945-8959.14.3.389>
- Petticrew, M. y Roberts, H. (2006). Systematic reviews in the social sciences. A practical guide. Oxford: Blackwell Publishing.

- Pillow, B. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development*, 3, 779-792.
- Quinn P. (2010). Born to Categorize. In U. Goswami, (Series Ed.). *Wiley-Blackwell handbook of childhood cognitive development.*, New York: Wiley (2nd ed, p. 129). doi.org/10.1111/1467-8624.00438
- Ricco, R. B. (2015). The Development of Reasoning. *Handbook of Child Psychology and Developmental Science*, 1–52. <https://doi.org/10.1002/9781118963418.CHILDPSY213>
- Rhodes, M., I, S. A., y Brickman, D. (2008). Developmental Changes in the Consideration of Sample Diversity in Inductive Reasoning. *Journal Of Cognition y Development*, 9(1), 112-143. <https://doi.org/10.1080/15248370701836626dien>
- Rhodes, M., y Liebensohn, P. (2015). Continuity and change in the development of category-based induction: The test case of diversity-based reasoning. *Cognitive Psychology*, 74. <https://doi.org/10.1016/j.cogpsych.2015.07.003>
- Ruffman, T., Taumoepeau, M., y Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, 30, 87–104, <https://doi.org/10.1111/j.2044-835X.2011.02045.x>
- Saffran, J. R., y Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, 69, 181–208. <http://dx.doi.org/10.1146/annurev-psych-122216-011805>.
- Samarapungavan, A., Vosniadou, S., y Brewer, W. F. (1996). *Mental models of the earth, sun, and moon: Indian children's cosmologies*. *Cognitive Development*, 11(4), 491–521. [https://doi.org/10.1016/s0885-2014\(96\)90015-5](https://doi.org/10.1016/s0885-2014(96)90015-5)
- Samarapungavan, A. Mantzicopoulos P. y Patrick H. (2008). Learning science through inquiry in kindergarten. *Science Education*. 92, 5, 868-908. doi 10.1002%2Fsce.20275 Schulz, L. (2012). The origins of inquiry: inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, 16, 7, 382-389. <https://doi.org/10.1016/j.tics.2012.06.004>
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32(1), 102–119. <https://doi.org/10.1037/0012-1649.32.1.102>

- Schulz, L. (2012). The origins of inquiry: inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, 16, 7, 382-389. [dx.doi.org/10.1016/j.tics.2012.06.004](https://doi.org/10.1016/j.tics.2012.06.004)
- Schulz, L., Goodman, N., Tenenbaum, J. y Jenkins, A. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, 109(2), 211-223. <https://doi.org/10.1016/j.cognition.2008.07.017>
- Schulz, L. E., y Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, 77(2), 427-442. <https://pubmed.ncbi.nlm.nih.gov/16611182>
- Schwichow, M., Croker, S., Zimmerman, C., Hoffler, T. y Hartig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 37. [doi-org.ezproxy.unal.edu.co/10.1016/j.dr.2015.12.001](https://doi.org/10.1016/j.dr.2015.12.001)
- Setoh, P., Wu, D., Baillargeon, R., y Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences, USA*, 110 (40), 15937-15942. <https://doi.org/10.1073/pnas.1314075110>
- Shafto, P., Kemp, C., Bonawitz, E. B., Coley, J. D., y Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition*, 109, 175-192. <https://doi.org/10.1016/j.cognition.2008.07.006>
- Strand-Cary, M., y Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, 23(4), 488-511. <https://doi.org/10.1016/j.cogdev.2008.09.005>
- Siler, S. A., Klahr, D., y Price, N. (2012). Investigating the mechanisms of learning from a constrained preparation for future learning activity. *Instructional Science*, 41(1), 191-216. <https://doi.org/10.1007/s11251-012-9224-7>
- Simon, H. A., y Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Potomac, MD: Erlbaum.
- Singh, M. (2021). Acquisition of 21st Century Skills Through STEAM Education. *Academia Letters*. <https://doi.org/10.20935/AL712>
- Shipley, E. y Shepperson, B. (2006). Test Sample Selection by Preschool Children: Honoring Diversity. *Memory and Cognition*, 34(7), 1444-1451, doi.org/10.3758/BF03195909

- Shtulman, A., y Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, 32 (6), 1049–1062. <https://doi.org/10.1080/03640210801897864>
- Shye, S. (1988). Inductive and Deductive Reasoning: A Structural Reanalysis of Ability Tests. *Journal Of Applied Psychology*, 73(2), 308-311
- Slone, L.K., Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks, *Cognition*, Volume 178, 2018, Pages 92-102, ISSN 0010-0277, <https://doi.org/10.1016/j.cognition.2018.05.016>.
- Sloman, S. A. (2005) *Causal models: how people think about the world and its alternatives*. Oxford University Press, Inc. : New York
- Sloutsky, V. M., Wei D., Fisher, A. y Kloos, H. (2015). Conceptual influences on induction: A case for a late onset, *Cognitive Psychology*, 82, 1-31. [dx.doi.org/10.1016/j.cogpsych.2015.08.005](https://doi.org/10.1016/j.cogpsych.2015.08.005)
- Sobel, D.M. y Buchanan D.W. (2009). Bridging the gap: Causality-at-a-distance in children's categorization and inferences about internal properties. *Cognitive Development*, 24, 274–283 <https://doi.org/10.1016/j.cogdev.2009.03.003>
- Sobel, D.M., Erbb, C.D., Tassina, T. y Skolnick Weisberg D. (2017). The Development of Diagnostic Inference About Uncertain Causes. *Journal of Cognition and Development*, Vol. 18, No. 5, 556–576 <https://doi.org/10.1080/15248372.2017.1387117>
- Sobel, D. M., y Legare, C. (2014). Causal learning in children. *WIRE-Cognitive Science*, 5, 413–427. <https://doi.org/10.1002/wcs.1291>
- Sternberg , R. (Ed.) (1984). *Mechanisms of cognitive development*. New York: Freeman
- Sternberg, R. (2010). Individual Differences in Cognitive Development. En U. Goswami, (Series Ed.). *Wiley-Blackwell handbook of childhood cognitive development*. New York: Wiley.
- Sternberg, R., y Gardner, M. (1983). Unities in inductive reasoning. *Journal Of Experimental Psychology: General*, 112(1), 80-116. <https://doi.org/10.1037/0096-3445.112.1.80>
- Sternberg, R. y Prezt, J. (Ed.) (2005). *Cognition and Intelligence: Identifying the Mechanisms of the Mind*. New York: Cambridge University Press.
- Sternberg, R. J., Sternberg, K., y Mio, J. S. (2012). *Cognitive psychology*. Australia: Wadsworth/Cengage Learning.

- Sü, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., y Schulze, R. (2002). Working-memory capacity explains reasoning ability-and a little bit more. *Intelligence -Norwood-Mutidisciplinary Journal*, (3), 261- 276. <https://www.sciencedirect.com/science/article/pii/S0160289601001003>
- Sutherland, S. L. y Cimpian, A. (2017). Inductive generalization relies on category representations. *Psychonomic Bulletin y Review*, 24(2), 632-636. <https://doi.org/10.3758/s13423-015-0951-z>
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., y Goodman, N. D. (2011). *How to Grow a Mind: Statistics, Structure, and Abstraction*. *Science*, 331(6022), 1279–1285. <https://doi.org/10.1126/science.1192788>
- Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., y Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, 332, 1054–1058. <https://doi.org/10.1126/science.1196404>
- Tummeltshammer, K.S., y Kirkham, N.Z. (2013). Learning to look: probabilistic variation and noise guide infants' eye movements. *Developmental Science*, 16, 760–771. <https://doi.org/10.1111/desc.12064>
- Tomasello, M. (2010). Language Development. In U. Goswami, (Series Ed.). *Wiley-Blackwell handbook of childhood cognitive development*. New York : Wiley.
- Tummeltshammer, K., Amso, D., French, R. M., y Kirkham, N. Z. (2016). Across space and time: infants learn from backward and forward visual statistics. *Developmental Science*, 1–9. doi.org/10.1111/desc.12474.
- Thurstone, L. L. y Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.
- Turiel, E. (2012). Moral reasoning, cultural practices, and social inequalities. *Innovación Educativa*, 12 (59), pp 17-32. www.scielo.org.mx/pdf/ie/v12n59/v12n59a3.pdf
- UNESCO. (2005). *Towards knowledge societies*. Paris: United Nations Educational, Scientific, and Cultural Organization.
- UNESCO (2015). *Replantear la educación: Hacia un bien común mundial?*. Disponible en: (www.unesco.org/open-access/terms-use-ccbysa-sp)

- UNESCO. (2017) *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*. París: United Nations Educational, Scientific and Cultural Organization. <http://unesdoc.unesco.org/images/0025/002534/253479e.pdf>
- Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *Journal Of Mathematical Psychology*, 6058-71. <https://doi.org/10.1016/j.jmp.2014.06.004>
- van Laar, E., Deursen, A., Van Dijk, J. y Haan, J. (2020). Determinants of 21st-Century Skills and 21st-Century Digital Skills for Workers: A Systematic Literature Review. *SAGE Open*. 10. 1-14. 10.1177/2158244019900176.
- Vega E. (2018). *The Wisdom of our Native American Tribes: Advanced Math, Science and Culture for the Future*. IEEE Integrated STEM Conference (ISEC). www.linkedin.com/in/ernesto-vega-janica-50231319
- Vygotsky, L. S. (1988). *El desarrollo de los procesos psicológicos superiores*. Barcelona, España: Crítica. (Año de publicación del original: 1931)
- Waismeyer, A., y Meltzoff, A. N. (2017). Learning to make things happen: Infants' observational learning of social and physical causal events. *Journal Of Experimental Child Psychology*, 58. <https://doi.org/10.1016/j.jecp.2017.04.018>
- Waismeyer, A., Meltzoff, A. N., y Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science*, 18, 175–182. <https://doi.org/10.1111/desc.12208>
- Washer, P. (2007). *Revisiting Key Skills: A Practical Framework for Higher Education*. *Quality in Higher Education*, 13(1), 57–67. <https://doi.org/10.1080/13538320701272755>
- Watson, G., y Glaser, E. M. (1964). *Manual for the Watson-Glaser critical thinking appraisal*. New York: Harcourt, Brace, Jovanovich
- Waxman, S. (2010). Early Word-Learning and Conceptual Development: Everything Had a Name, and Each Name Gave Birth to a New Thought. In U. Goswami, (Series Ed.). *Wiley-Blackwell handbook of childhood cognitive development*. (2nd ed.). New York: Wiley.
- Waldmann, M. y Hagmayer, Y. (2006). Categories and causality: The neglected direction. *Cognitive Psychology*, 53, 27-58. <https://doi.org/10.1016/j.cogpsych.2006.01.001>

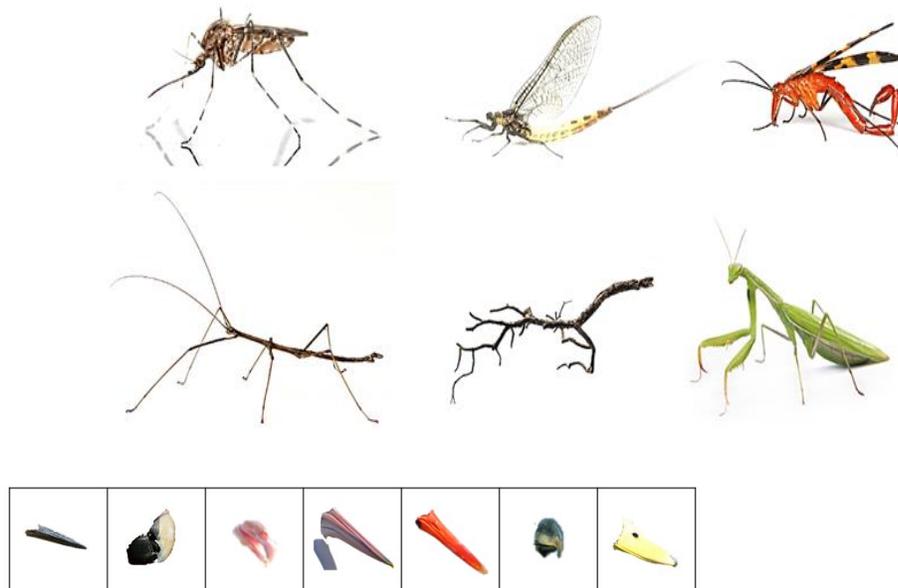
- Wellman, H. M. (2014). *Making Minds: How Theory of Mind Develops*. Oxford University Press : New York
- Wild y Pfannkuch (1999), Statistical Thinking in Empirical Enquiry, *International Statistical Review*, 67, pp. 223 –265. doi.org/10.1111/j.1751-5823.1999.tb00442.x
- Wilkening, F., y Cacchione, T. (2010). Children's intuitive physics. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 473–496). Wiley-Blackwell.
- World Economic Forum. (2020). *Schools of the Future: Defining New Models of Education for the Fourth Industrial Revolution*, 2020. [WEF Schools of the Future Report 2019.pdf \(weforum.org\)](https://www.weforum.org/reports/schools-of-the-future-report-2019)
- World Economic Forum. (2021). *Building A Common Language for Skills at Work: A Global Taxonomy*. <https://www.weforum.org/reports/building-a-common-language-for-skills-at-work-a-global-taxonomy>
- Wu R, Gopnik A, Richardson DC, Kirkham NZ. (2011). Infants learn about objects from statistics and people. *Dev. Psychol.* 47(5):1220–29. [https://doi.org/ 10.1037/a0024023](https://doi.org/10.1037/a0024023)
- Xu, F. y Garcia, V. (2008). Intuitive Statistics by 8-Month-Old Infants. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 105 (13), 5012-5015. <https://doi.org/10.1073/pnas.0704450105>
- Xu, F., y Griffiths, T. L. (2011). Probabilistic models of cognitive development: towards a rational constructivist approach to the study of learning and development. *Cognition*, 120 (3), 299–301 doi.org/10.1016/j.cognition.2011.06.008
- Xu, F. y Kushnir, T. (Eds.). (2012). *Rational Constructivism in Cognitive Development*. Waltham, MA: Academic Press.
- Zheng, D., Shu G., y Wong, K. W. (2018). Global Engineering. *Education Conference (EDUCON) - Santa Cruz de Tenerife, Islas Canarias, España*. <https://doi.org/10.1109%2FEDUCON.2018.8363268>
- Zhu, L., y Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, 98(3), 287–308. <https://doi.org/10.1016/j.cognition.2004.12.003>
- Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review*, (20), 99-149. <https://doi.org/10.1006/drev.1999.0497>

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 2, 172-223. <https://doi.org/10.1016/j.dr.2006.12.001>

Anexo A. Imágenes de actividades de Familiarización por Clasificación del IMERIN-v1.



ANEXO B. Imágenes de secuencias en las actividades de categorización del IMERIN-v1.

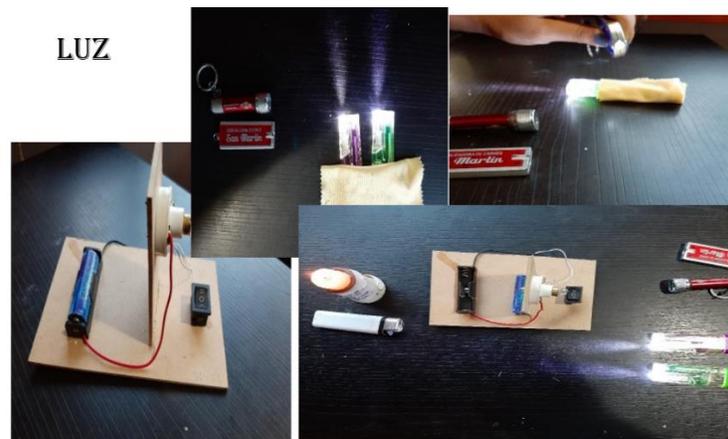




ANEXO C. Imágenes de secuencias de actividades de generalización por categorización del IMERIN-v1.



ANEXO D. Ejemplos de actividades de RI en la modalidad intervención, del IMERIN- v1.



ASTROS- SATÉLITES



ANEXO E. Ejemplos de actividades de RI en la modalidad elección de evidencia y terminación de secuencias del IMERIN- v1.

PROFESIONES

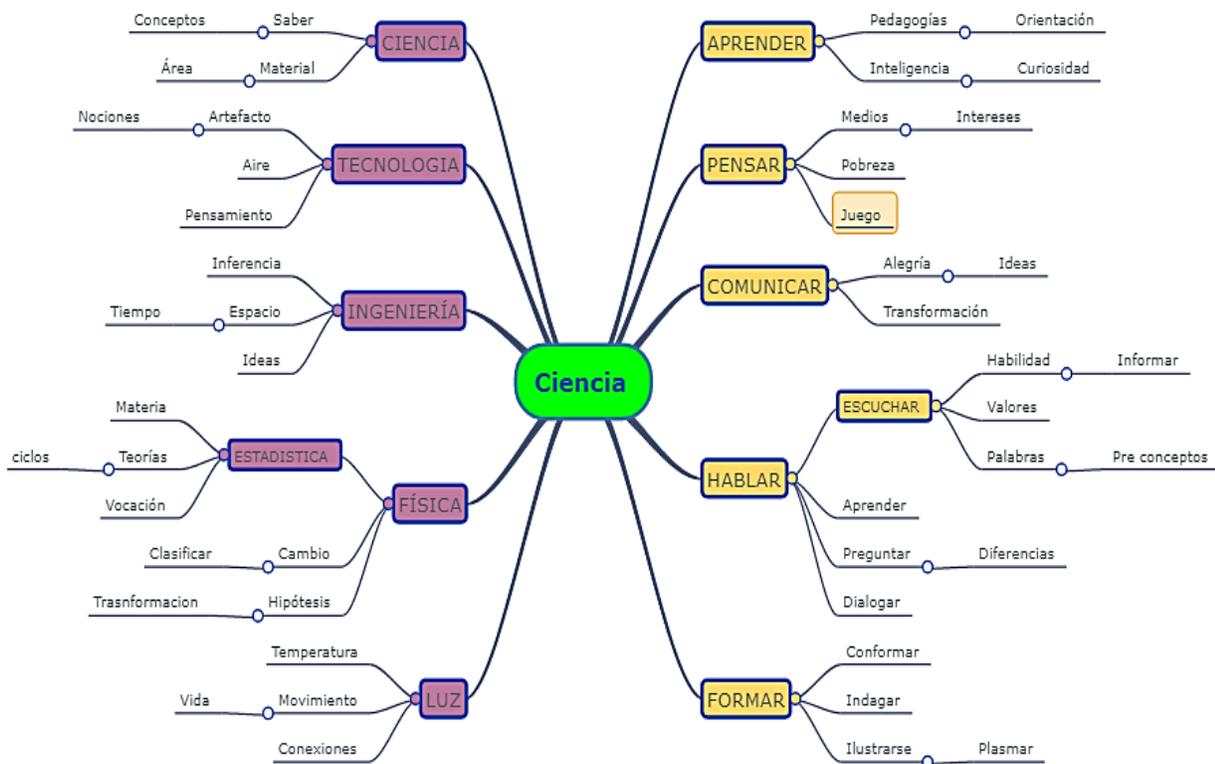


MONOS



ANEXO F. Estudio de redes semánticas

Con la participación de profesoras de Preescolar, profesoras de Primaria y madres de niños de 4 a 5 años, desescolarizados durante 2020 y madres de niños vinculados al jardín. Este tipo de estudio se justifica cuando hay evidencia confiable de que, en los grupos humanos vinculados a la investigación, se usan diferentes palabras para nombrar el concepto de interés. Dado que los fundamentos de las actividades de RI son los principios teóricos, términos técnicos y alusivos al lenguaje especializado de los dominios STEM, fue necesario estudiar el significado connotativo o la connotación que tienen algunas palabras en los contextos cotidianos de las familias de los niños participantes.



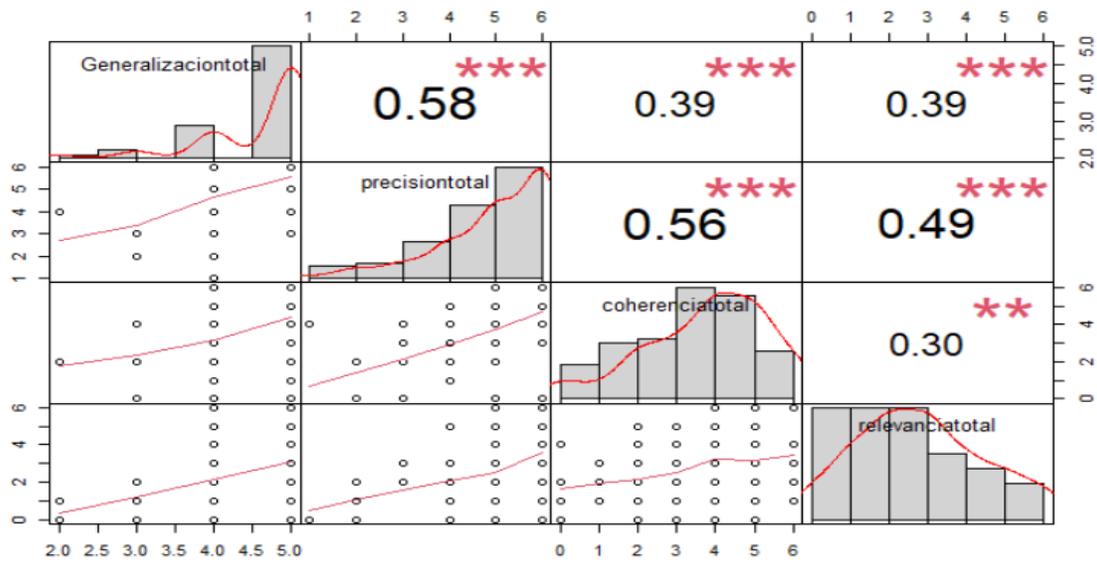
<p>Identifica plastilina de colores primarios y la intervención de un agente que las manipula</p>	<p>agente cambia propiedades evidentes de la plastilina</p>	<p>Precisión: Suponer que para tener un nuevo color – café- se deben mezclar nuevas plastilinas de colores diferentes a los mezclados anteriormente, o usar las bolitas que quedaron después de mezclar los primarios, o mezclar las tres bolitas, en vez de dos (como en fase 1)</p>
<p>Identifica el cambio en las propiedades como un efecto probablemente causado por la</p>	<p>Predice coherentemente posibles efectos de una nueva acción sobre</p>	<p>Coherencia: Predecir que podrían usarse los mismos materiales que</p>

Anexo I. Resultados del estudio de concordancia en el juicio de expertos, con coeficiente ω de Kendall

Coeficiente W de Kendall		Categoría de Juicio		
		Pertinencia	Suficiencia	Redacción
Dimensiones juzgadas en cada actividad -Precisión - Coherencia - Relevancia	Actividad 1	0.12	0.25	1
	Actividad 2	0.25	1	1
	Actividad 3	1	0.25	1
	Actividad 5	0.25	1	1
	Actividad 6	1	0.25	1
	Actividad 7	0,37	0,37	1
	Actividad 8	0.25	0.25	1
	Actividad 9	0	0.25	0,33

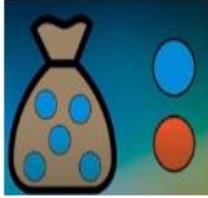
Nota: uno de los jueces solo juzgó 2 de las 3 dimensiones, por lo que la información del índice de concordancia es limitada.

Anexo J. Gráfico de la matriz de correlaciones policóricas – acierto en cada una de las 4 dimensiones evaluadas. Análisis de consistencia interna de la prueba IMERIN-P

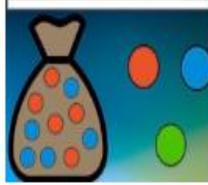


Anexo K. Ejemplo de Actividad retirada del IMERIN-P – Concepto Jueces

TAREA No. 5: CERTEZA VS INCERTIDUMBRE



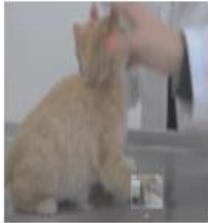
¿la bolita roja salió de esa bolsa?... podemos estar seguros?



¿la bolita verde salió de esa bolsa?... tal vez era la única que había y salió...

¿Si se puede ver el contenido de la bolsa, se puede estar seguro de cuáles bolitas salieron de ésta?... y si no vemos el contenido de la bolsa, podemos adivinar el color de la bolita que saldrá?
¿De qué color será la bolita que saque alguien que no ha visto lo que contiene?

FASE 2



¿la veterinaria está segura de que el purgante alentar al gatito?



¿Por qué la médica le está poniendo esos cables al sr?



¿Podrán las astronautas arreglar una nave espacial o decirnos si algo le está fallando?

Anexo L. Análisis de varianza Puntaje total y cada uno de los ítems del IMERIN, prueba Kruskal-Wallis (es una prueba no paramétrica de comparación de medianas de muestras independientes)

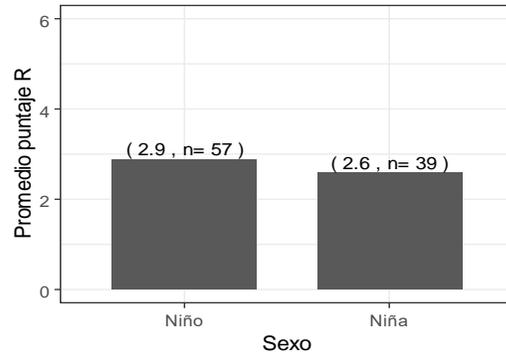
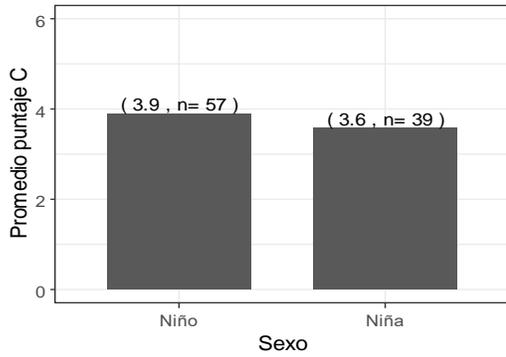
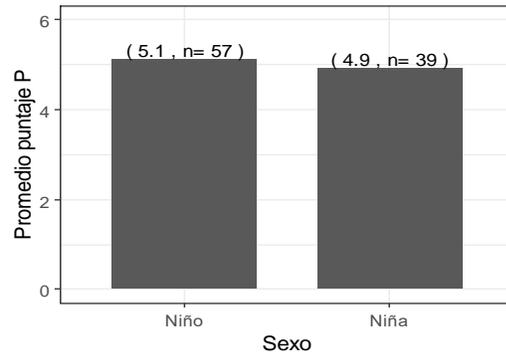
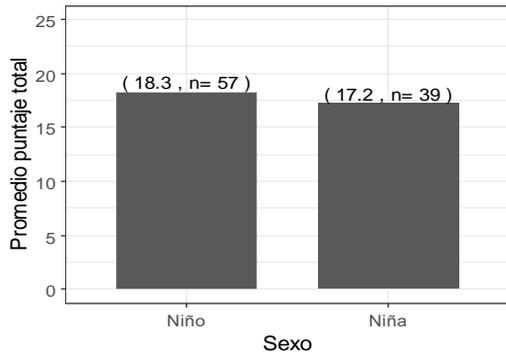
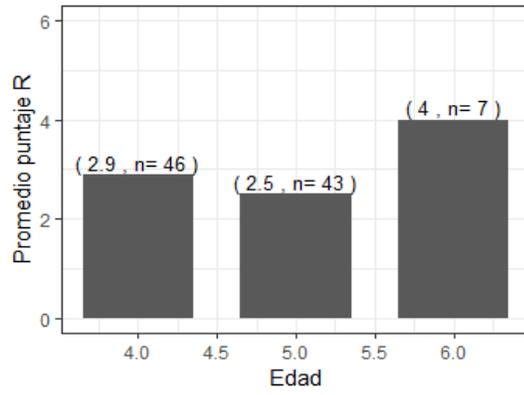
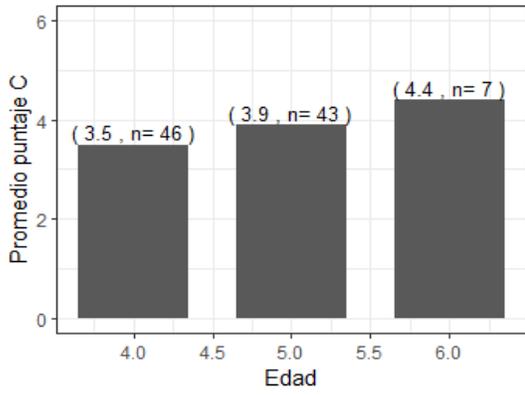
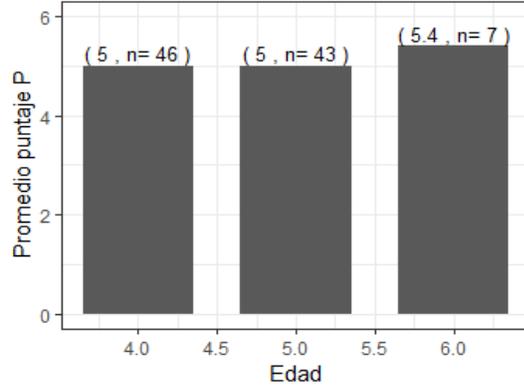
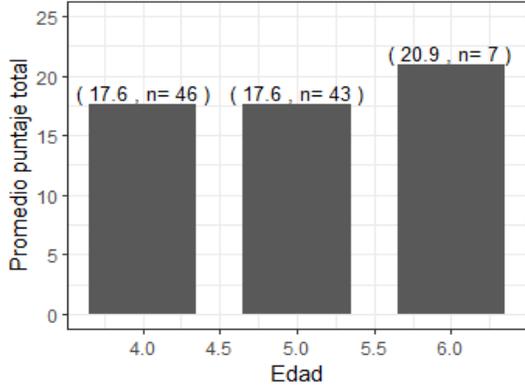
Ítem	<i>p valor</i>
CP1	7.155e-07*
CC1	1.245e-07*
CR1	0.007711
CP2	0.03037
CC2	2.98e-06*
CR2	0.0006946
EP1	0.0146
EC1	0.001239
ER1	3.824e-06*
EP2	0.000624
EC2	0.0001267
ER2	2.545e-07*
TP1	1.558e-08*
TC1	1.858e-07*
TR1	6.148e-07*
TP2	1.103e-06*
TC2	0.02706
TR2	3.962e-06*

Nota: C= dominio Ciencias; E= dominio Estadística; T= dominio Tecnología; P= dimensión precisión; C= dimensión coherencia; R= dimensión relevancia *=notación estadística, especifica la cantidad de ceros (0) a la izquierda del primer número del dato.

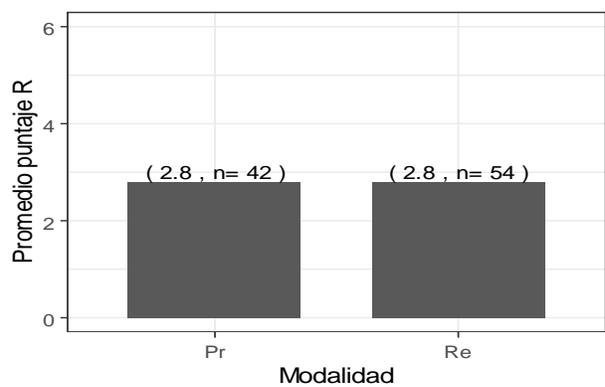
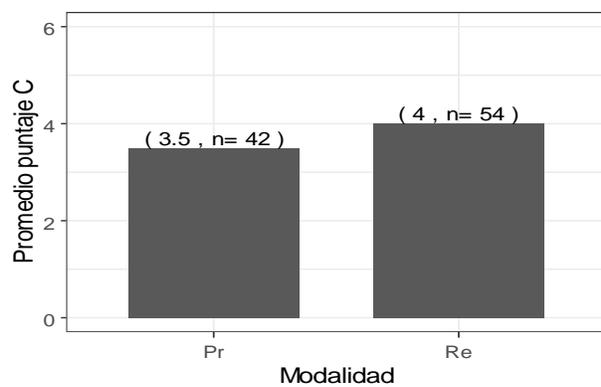
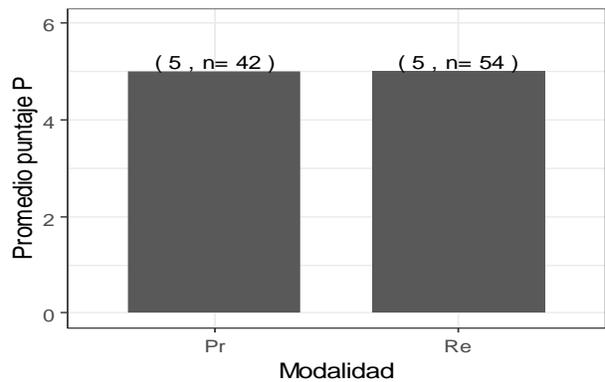
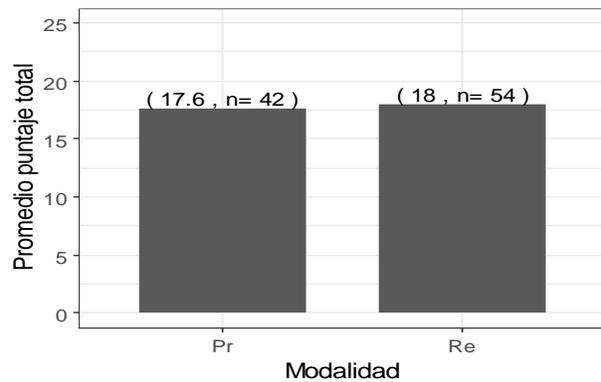
Anexo M. Correlaciones Sperman entre puntajes en el ítem y edad del participante (N=24)

		item	Media	Desv.	Rho
Ciencias explicación	Generalización	1	0,93	0,2	-0.04
	Precisión	2	0,85	0,3	-0.007
	Coherencia	3	0,78	0,4	-0.01
	Relevancia	4	0,51	0,4	-0.005
Ciencias predicción	Generalización	5	0,96	0,1	0.18
	Precisión	6	0,92	0,2	0.02
	Coherencia	7	0,76	0,4	0.12
	Relevancia	8	0,53	0,4	-0.01
Estadística explicación	Generalización	9	0,97	0,1	-0.1
	Precisión	10	0,84	0,3	-0.006
	Coherencia	11	0,52	0,4	0.11
	Relevancia	12	0,44	0,4	-0.01
Estadística predicción	Generalización	13	0,91	0,2	0.03
	Precisión	14	0,83	0,3	-0.0005
	Coherencia	15	0,55	0,4	0.09
	Relevancia	16	0,45	0,4	-0.09
Tecnología explicación	Generalización	17	0,84	0,3	0.02
	Precisión	18	0,72	0,4	0.02
	Coherencia	19	0,41	0,4	0.06
	Relevancia	20	0,29	0,4	-0.001
Tecnología predicción	Generalización	21	0,95	0,1	0.11
	Precisión	22	0,82	0,3	0.17
	Coherencia	23	0,72	0,4	0.15
	Relevancia	24	0,56	0,4	0.15

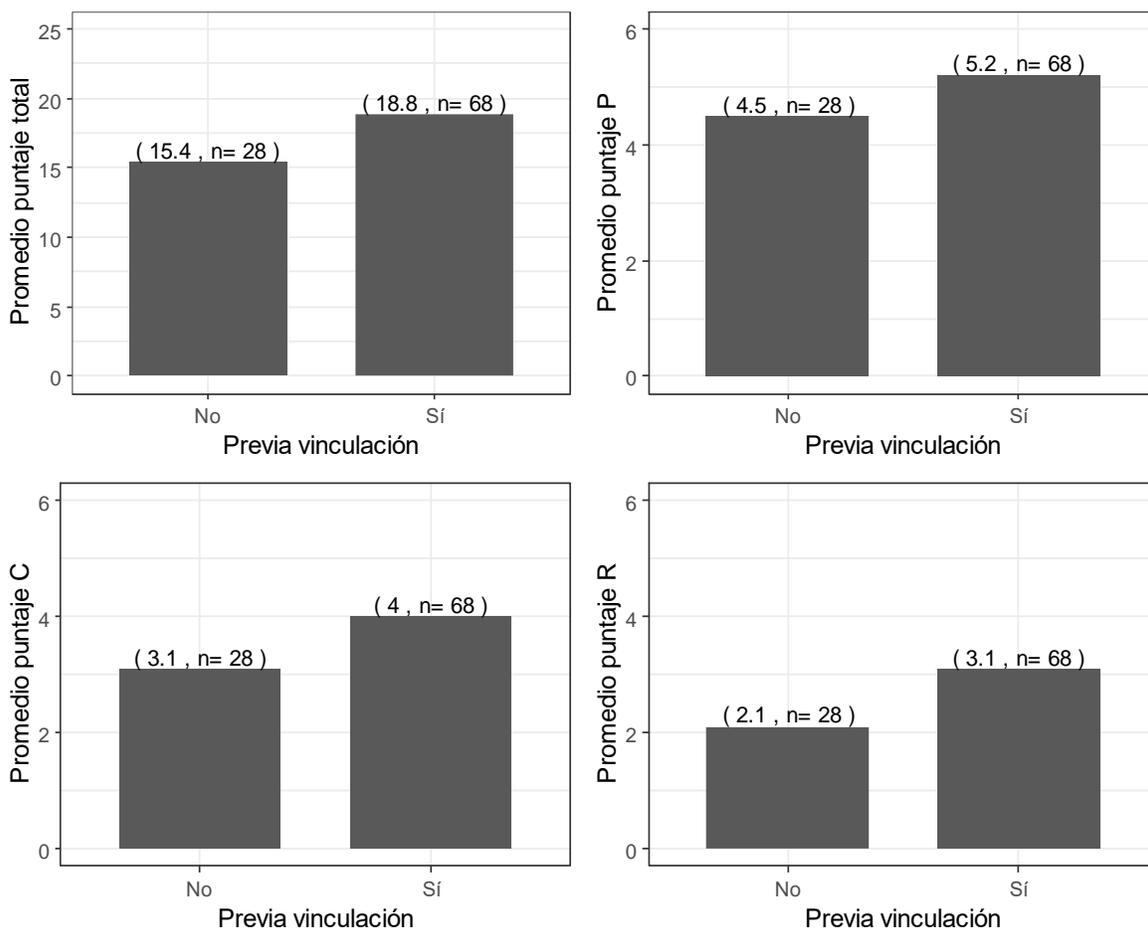
Nota: correlación significativa al 0%, $p \leq 0.010$



Anexo O. Gráficos de puntuaciones promedio en las subescalas del IMERIN discriminados por modalidad de la aplicación



Anexo P. Gráficos de puntuaciones promedio en las subescalas del IMERIN discriminados por condición de vínculo institucional previo



Anexo Q. Ítems de la encuesta sobre prácticas de socialización de la cognición- Dirigida a madres y padres de niños participantes

1. Propongo conversaciones sobre posibles soluciones a problemas que afectan a todas las personas del mundo, por ejemplo: la contaminación, o el cambio climático
2. Invento actividades para comparar y explicar las similitudes y diferencias entre algunas clases de objetos, máquinas, o seres vivos que se encuentran en nuestro entorno
3. Doy ejemplos para explicarle que las personas tenemos la capacidad de pensar y generar conocimientos, desde que somos niñas y niños
4. He aprendido sobre los temas o situaciones que le generan curiosidad, preguntas o interés.
5. Brindo información para que comprenda las causas por las que funcionan los aparatos eléctricos, el control remoto, o de otros funcionamientos que ocurren por causas invisibles.
6. ¿Enseña a su hija sobre algún tema que a usted le interesa?. Si su anterior respuesta fue SI, por favor indique de que área es el tema.
 - 1Agricultura, medio ambiente, ecología, producción agropecuaria
 - 2Artes, Deportes, Religión, Cultura
 - 3Ciencias, Ingeniería, Matemáticas, Tecnología
 - 4Comercio, Economía, Finanzas, Negocios
7. Realizamos actividades para descubrir las causas de eventos que podrían generarle curiosidad, por ejemplo: mezclar dos colores para obtener un tercer color diferente
8. Pregunto su opinión o creencias sobre hechos ocurridos recientemente y que son comentados en las noticias, o por otros medios
9. He confirmado que la información que cree que es verdadera, realmente si lo es, por ejemplo: información sobre cómo evitar enfermedades causadas por virus y bacterias
10. Dedico tiempo a aclarar si ha aprendido ideas adecuadas sobre los temas que yo, o alguien más le ha enseñado.
11. Invento historias poco creíbles sobre hechos que ya conoce, para generarle dudas y que se motive a verificar lo que estoy diciendo
12. Cuando hablamos, uso los nombres y palabras exactas para mencionar los objetos, hechos, personajes o sujetos a los que me refiero.
13. Propongo conversaciones que le faciliten aprender el significado de palabras relacionadas con temas de su interés.

14. Le cuento sobre los nuevos aprendizajes que yo he logrado y le pregunto sobre los suyos.
15. Hago lecturas en voz alta sobre temas o áreas de conocimiento que me gustaría que comprenda, le interesen y aprenda
16. Expreso mis expectativas o intenciones sobre las clases de conocimientos que me gustaría que aprendiera en la escuela.
17. ¿Usted cree que tanto las niñas como los niños tienen las mismas oportunidades para aprender conocimientos, en cualquier área del saber o sobre cualquier tema? Por favor explique su respuesta:
18. En cuál área del conocimiento le gustaría que su hija se motive a aprender a lo largo de su infancia? Por ejemplo:
- 1 Agricultura, medio ambiente, ecología, producción agropecuaria
 - 2 Artes, Deportes, Religión, Cultura
 - 3 Ciencias, Ingeniería, Matemáticas, Tecnología
 - 4 Comercio, Economía, Finanzas, Negocios