

No. **61**

Marzo 2019

ISSN 2215 – 7816 (En línea)

Documentos de Trabajo

Escuela de Gobierno Alberto Lleras Camargo

Improving the Measurement of
Children's Mental Health Problems in
Colombia with Item Response Theory

Jorge Cuartas

Serie Documentos de Trabajo 2019

Edición No. 61

ISSN 2215-7816 (En línea)

Edición digital

Marzo de 2019

© 2019 Universidad de los Andes, Escuela de Gobierno Alberto Lleras Camargo

Carrera 1 No. 19 -27, Bloque AU

Bogotá, D.C., Colombia

Teléfono: 3394949, ext. 2073

escueladegobierno@uniandes.edu.co

<http://egob.uniandes.edu.co>

Director Escuela de Gobierno Alberto Lleras Camargo

Eduardo Pizano de Narváez

Autor

Jorge Cuartas

Jefe de Mercadeo y Comunicaciones, Escuela de Gobierno Alberto Lleras Camargo

Camilo Andrés Torres Gutiérrez

Gestora Editorial, Escuela de Gobierno Alberto Lleras Camargo

Angélica María Cantor Ortiz

Gestor de Comunicaciones, Escuela de Gobierno Alberto Lleras Camargo

Camilo Andrés Ayala Monje

El contenido de la presente publicación se encuentra protegido por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por tanto su utilización, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso, digital o en cualquier formato conocido o por conocer, se encuentran prohibidos, y solo serán lícitos en la medida en que cuente con la autorización previa y expresa por escrito del autor o titular. Las limitaciones y excepciones al Derecho de Autor solo serán aplicables en la medida en se den dentro de los denominados Usos Honrados (Fair Use); estén previa y expresamente establecidas; no causen un grave e injustificado perjuicio a los intereses legítimos del autor o titular; y no atenten contra la normal explotación de la obra.

Improving the Measurement of Children's Mental Health Problems in Colombia with Item Response Theory*

By Jorge Cuartas[†]

* I would like to thank Andrew Ho for all his support in the process of writing this manuscript, and to Catalina Rey-Guerra, Ben West, Philipp Hessel, and the members of S061 at the Harvard Graduate School of Education for their comments. All data for this study were taken from publicly available data sources. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

[†] Harvard Graduate School of Education, Harvard University, 13 Appian Way, Cambridge, MA, The United States of America. E-mail address: jcuartas@g.harvard.edu.

Abstract

The present study examines the psychometric properties of the mental health scale for children used in the 2015 Colombian Mental Health Survey. A nationally representative sample of 2,727 children is used ($M_{age} = 8.99$; $range = 7 - 11$) that includes reports from their main caregivers about 26 mental health problem symptoms from the Reporting Questionnaire for Children (RQC), the Child Behavior Checklist (CBCL) and the Brief Screening and Diagnostic Questionnaire (CBTD). Classical test theory and factor analysis were conducted to analyze the classical location and information of each item, along with the dimensionality, reliability, and convergent validity of the scale. Item Response Theory (IRT) was used in order to estimate theoretically invariant item parameters for location and information. Findings reveal that the mental health scale for children has adequate psychometric properties for its use in Colombia. Furthermore, IRT analyses reveals a set of items that maximize information and that may be used for future use when more efficiency is warranted.

Keywords: Item Response Theory, Test Information Function, Classical Test Theory, Children's Mental Health, Psychometrics.

Resumen

El presente artículo examina las propiedades psicométricas de la escala de salud mental para niños incluida en la Encuesta Nacional de Salud Mental 2015. Se utiliza una muestra representativa nacional de 2,727 niños ($M_{edad} = 8.99$; $rango = 7 - 11$), con reportes de sus cuidadores principales sobre 26 síntomas de problemas de salud mental tomados de los instrumentos Reporting Questionnaire for Children, Child Behavior Checklist, y el Brief Screening and Diagnostic Questionnaire. Se utilizaron teoría clásica de los test y análisis factorial para analizar los parámetros de localización e información clásicos de cada ítem, al igual que la dimensionalidad, confiabilidad y validez convergente de la escala. Se utilizó Teoría de Respuesta al Ítem (TRI) para estimar parámetros de localización e información teóricamente invariantes. Los resultados revelan que la escala tiene propiedades psicométricas adecuadas. Los resultados de TRI indican un conjunto de ítems que maximizan la información de la prueba y que podrían utilizarse en una administración futura que requiera mayor eficiencia.

Palabras clave: Teoría de Respuesta al Ítem, función de información del test, teoría clásicas de los test, salud mental infantil, psicometría

Table of content

| | |
|---|----|
| 1. Introduction..... | 5 |
| 2. Methods..... | 7 |
| 2.1 Participants..... | 7 |
| 2.2 Instruments..... | 7 |
| 2.3 Statistical analysis..... | 8 |
| 3. Results..... | 10 |
| 3.1 Classical test theory and dimensionality..... | 10 |
| 3.2 IRT and scale information..... | 12 |
| 4. Discussion..... | 16 |
| 5. Conclusion..... | 17 |
| 6. References..... | 18 |
| 7. Appendix..... | 21 |

1. Introduction

Recent estimates suggest that between 10 to 20 percent of children and adolescents in low- and middle-income countries (LMICs) suffer from mental health problems (Erskine et al., 2017; Kieling et al., 2011). Despite this high prevalence, the mental health needs of children living in LMICs are often unattended for lack of funding, political indifference, or lack of qualified clinicians (Kieling et al., 2011). Children's mental health problems are particularly prevalent in conflict-affected countries, but these countries may also be less capable of monitoring their children's mental health needs (Dimitry, 2012). In this context, it is fundamental to develop efficient and reliable instruments to assess and monitor children's mental health problems to make visible the problem and promote efforts aimed at reducing it.

Colombia has suffered from more than 50 years of civil conflict, leaving more than 1.4 million children and adolescents as direct victims (Red Nacional de Información, 2018). Until 2015, the country did not have information about the national prevalence of mental health problems in children, despite evidence suggesting it was high in specific regions (OIM, UNICEF, & ICBF, 2013). In 2015, Colombia has implemented its first nationally representative mental health survey for children aged 7 to 11 years (ENSM for its acronym in Spanish; Ministerio de Salud & Colciencias, 2015), which included 24 items selected from different scales by subject-matter experts to assess mental health problems (Rodriguez et al., 2016). Even though the content validity of these items was discussed (see Ministerio de Salud & Colciencias, 2015), the psychometric properties of this scale, which comprises items from different sources, has not been assessed. The present study seeks to fill this gap by analyzing the mental health problems' scale dimensionality, reliability, and convergent validity. Moreover, this study seeks to analyze evidence on the information each item provides, in order to define whether a more efficient scale (i.e., shorter and with a high level of information) is feasible to facilitate a future implementation to monitor children's mental health problems.

Even though Classical Test Theory (CTT) is the most widely used framework to analyze the psychometric properties of test scores, Item Response Theory (IRT) offers unique features to improve the efficiency of a scale. Contrary to CTT, where item statistics (e.g., percentage of items correct, item-test correlation, measures of reliability) are population-dependent (Lord, Novick, & Birnbaum, 1968), IRT estimates item characteristics that are assumed as invariant across

populations, occasions, and independent of other items embedded in the test or questionnaire (Brennan, National Council on Measurement in Education, & American Council on Education, 2006). In particular, IRT estimates item discrimination (referring to the extent to which the item is capable of distinguishing between individuals with different levels of the latent trait) and item location (referring to the level of the latent trait where the scale is most reliable and precise in distinguishing between individuals) (Embretson & Reise, 2013).

Another feature that makes IRT stand out is that it recognizes that a scale will be more reliable and precise in distinguishing between individuals at a certain segment of the latent continuum, whereas CTT assumes a single homogenous estimate of reliability (Brennan et al., 2006). Taking together, IRT provides information that can be used for the design of most efficient scales, selecting items that provide higher levels of information at the levels of the latent trait of interest (Jessen, Ho, Corrales, Yueh, & Shin, 2018). In turn, it is possible to obtain a shorter and easier to implement scale, as well as a set of items that allow a more reliable and targeted measure.

In the case of the mental health scale for children used in the ENSM, it is unclear whether the 26 items selected by subject-matter experts represent a single underlying dimension, hypothesized to be mental health problems. Moreover, it is unclear whether a future implementation, using a shorter but high-informing scale, is possible. These issues are critical not only to reduce the time and resources used at measuring children's mental health problems in Colombia in its post-conflict situation but also to do so with precision to thoroughly inform prevention and attention efforts across the country. This study contributes to these objectives by answering the following research questions:

1. Does the mental health scale used in the ENSM measure a single factor (i.e., mental health problems) as intended?
2. What are the psychometric properties of the mental health scale, according to CTT and IRT frameworks?
3. Is it possible to implement a more efficient (i.e., with fewer items and high precision) scale for children's mental health problems on a future occasion?

2. Methods

2.1 Participants

The ENSM is a nationally representative sample for non-institutionalized children aged 7 to 11 years, with representability for four regions (Atlantic, Western, Central, and Pacific) and Bogotá, and each of the 32 national departments. The ENSM sampling comprises a probabilistic, multistage sampling procedure, and the sample size was designed following findings from previous national studies (Rodríguez et al., 2016). The sample used in the present study includes 2,727 children, having complete information for all the cases included in the ENSM. According to the ENSM, using a sample of children that were on average 9 years old, more than half of them were girls, 19 percent of Colombian children aged 7 to 11 years belong to an ethnic minority, around 58 percent lived with their parent and 86 percent with their mother, 98 percent attended school, and 21 percent were considered poor according to a multidimensional poverty index (Alkire & Foster, 2011). Table 1 presents details. ENSM surveys were collected between January and May 2015.

Table 1. Sample characteristics ($N = 2,727$)

| Variable | M | SD | Min | Max |
|--|----------|-----------|------------|------------|
| Age | 8.99 | 1.41 | 7 | 11 |
| Sex (=1 if male) | 0.49 | 0.50 | 0 | 1 |
| Ethnic minority | 0.19 | 0.39 | 0 | 1 |
| Liveing with father | 0.58 | 0.49 | 0 | 1 |
| Lived with mother | 0.86 | 0.31 | 0 | 1 |
| Maternal education | | | | |
| Less than basic | 0.05 | 0.22 | 0 | 1 |
| Basic | 0.27 | 0.44 | 0 | 1 |
| Secondary | 0.52 | 0.49 | 0 | 1 |
| Superior | 0.17 | 0.37 | 0 | 1 |
| Attended school | 0.98 | 0.15 | 0 | 1 |
| Reported-health status | 4.02 | 0.99 | 0 | 5 |
| Multidimensional poor household | 0.21 | 0.41 | 0 | 1 |
| Region | | | | |
| Central | 0.24 | 0.42 | 0 | 1 |
| Atlantic | 0.24 | 0.42 | 0 | 1 |
| Bogotá D.C. | 0.14 | 0.35 | 0 | 1 |
| Western | 0.21 | 0.41 | 0 | 1 |
| Pacific | 0.17 | 0.37 | 0 | 1 |

Note: Averages using sample weights

Source: author

2.2 Instruments

The ENSM included a 26-item instrument to assess children's mental health problems. This instrument includes 10 items from the Reporting Questionnaire for Children (RQC; Giel et al.,

1981), and additional items from the Child Behavior Checklist (CBCL; Achenbach, 1999) the Brief-Screening Diagnostic Questionnaire (CBTD, for its acronym in Spanish; Caraveo y Anduaga, 2007), and “others based on the experience of the research groups [i.e., subject-matter experts that participated in the ENSM]” (Rodríguez et al., 2016, p. 15). The 26 items included in the scale, presented in Appendix 1, refer to yes (coded as 1) or no (coded as 0) questions, aimed at identifying diverse symptoms of mental health problems in children. One of the main caregivers reported children's symptoms: 80 percent were their mothers, 7.2 percent their fathers, and the remaining were other caregivers. Even though previous studies show that the RQC, CBCL, and CBTD have good psychometric properties for assessing children’s mental health problems (e.g., Ahmad et al., 2007; Castro, Billick, & Swank, 2016), to date the validity evidence on the internal structure (i.e., coherence) for the scale employed in the Colombian ENSM (which pooled items from different scales) has not been assessed.

2.3 Statistical analysis

To begin with, Classical Test Theory (CTT) statistics were estimated (Crocker & Algina, 1986; Novick, 1966; Traub & Rowley, 1991). To analyze the characteristics of each item, the percentage of affirmative answers were used as a classical item location estimate and the item-test correlation as a classical information estimate. For the overall scale, Cronbach’s alpha was estimated to examine the reliability (i.e., internal consistency) of the scale. Subsequently, factor analysis was used to fit a unidimensional model to the data and assess the dimensionality of the scale by analyzing the share of variance accounted for by the first factor (Thompson, 2004).

Furthermore, following the model presented in Equation 1, a two-parameter IRT model was fit to the data (Embretson & Reise, 2013; Lord et al., 1968). In the model, θ_p represents the latent score (i.e., mental health problems - symptoms) of each children p , which is standardized (i.e., mean of zero and standard deviation of 1). Additionally, a_i represents the discrimination parameter, which indicates how well an item can distinguish between children with slightly different levels of the latent variable, and it is alike to factor loadings in a confirmatory factor analysis where items are continuous. Particularly, the discrimination parameter shows that 1 unit increase in the latent variable θ produces an a increase in the log of the odds of answering affirmatively the item. Finally, b_i represents item location, which shows the level of the latent variable θ at which children have even odds of answering each item affirmatively. An advantage

that IRT has over CTT is that the former estimates parameters that are considered to be invariant to populations of items and individuals, whereas the later produce population-dependent parameters (Embretson & Reise, 2013; Traub & Rowley, 1991).

$$P_i(\theta_p) = \frac{1}{1 + \exp(-a_i(\theta_p - b_i))}; \theta_p \sim N(0,1) \quad (1)$$

After fitting the IRT model to the data, the test information function is estimated as the sum of all item's information functions, which are calculated through their discrimination parameters (a) and the product of the likelihoods of having an affirmative (P) or negative (Q) answer in the item (Equation 2).

$$I(\theta) = \sum_i I_i(\theta_p) = \sum_i a_i^2 P_i(\theta) Q_i(\theta) \quad (2)$$

Using the test information function, the conditional errors of measurement (SE) was computed as presented in Equation 3. Contrary to CTT, where there is a single standard error of measurement for the scale, the SE in IRT shows the estimated error at different locations of the scale, allowing to assess the level of imprecision in the measurement at different levels of θ .

$$SE(\hat{\theta}|\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (3)$$

Subsequently, items that provided less information were flagged for potential exclusion in a shortened version of the scale. Particularly, the information provided by the scale comprising all the original items (i.e., 26) was compared with a reduced scale with 21 items, 18 items, 14 items, and 11 items, removing items that provided less information in a step-wise fashion. To provide validity evidence based on correlation for the test scores calculated with different items, convergent validity evidence was analyzed, using information gathered using the Diagnostic Interview for Children (DISC-IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), which is an instrument that evaluated 32 common psychiatric diagnoses of children based on the Diagnostic and Statistical Manual – IV (DSM-IV; Bell, 1994). Convergent validity evidence was also examined through correlations between the total score and children's reported physical health (Aarons et al., 2008), exposure to bullying or discrimination (Cooke, Bowie, & Carrère, 2014), and exposure to selected adversities (e.g., Anda et al., 2006), such as exposure to community crime, parental separation, major sickness, and other stressful events, which have been widely linked to mental health problems during childhood. All analyses were conducted in Stata 15.1 (StataCorp, 2017).

3. Results

3.1 Classical test theory and dimensionality

The 26-item scale exhibits good reliability, with a Cronbach’s alpha (α) of 0.74, suggesting that 74 percent of observed score variance is accounted for by true score variance, according to CTT. As shown in Table 1, according to CTT estimates all the items have high location parameters (i.e., only a small proportion of children present the assessed symptoms, see also Figure A1 for histograms), whereas there is considerable variability in the amount of information that each item provides (according to item-test correlation), ranging from 0.20 (item 21, “Has the child needed to change school more than 3 times?”) to 0.58 (item 12, “Have you noticed that the child has difficulty making friends of his or her same age?”). Nonetheless, these parameters are population dependent, so in a different administration with a different sample they may vary.

Table 2. Classical test theory analysis ($N = 2,727; \alpha = 0.74$)

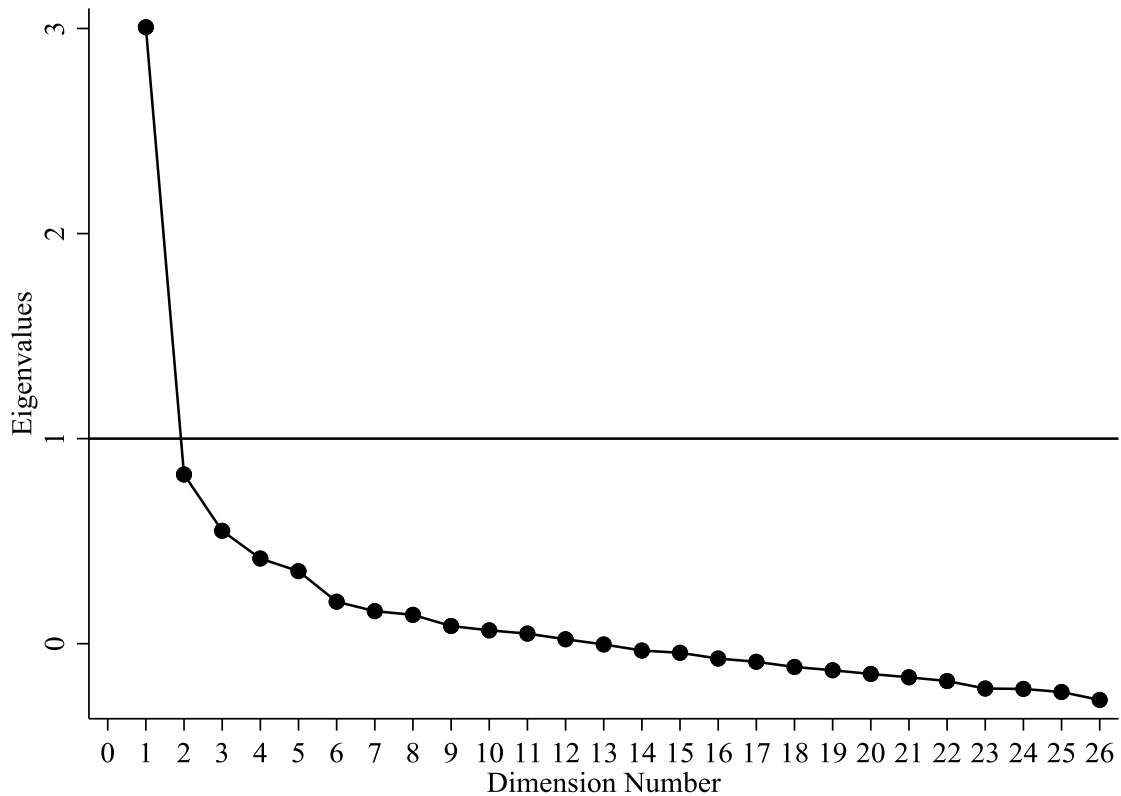
| | Classical location | Classical information |
|---------|---------------------------|------------------------------|
| | Affirmative (%) | Item-test correlation |
| Item 1 | 2.4 | 0.23 |
| Item 2 | 0.9 | 0.24 |
| Item 3 | 4.2 | 0.41 |
| Item 4 | 3.7 | 0.38 |
| Item 5 | 1.8 | 0.34 |
| Item 6 | 1.8 | 0.31 |
| Item 7 | 12.4 | 0.55 |
| Item 8 | 18.6 | 0.51 |
| Item 9 | 6.8 | 0.39 |
| Item 10 | 4.8 | 0.36 |
| Item 11 | 5.0 | 0.30 |
| Item 12 | 15.4 | 0.58 |
| Item 13 | 3.0 | 0.29 |
| Item 14 | 3.7 | 0.35 |
| Item 15 | 7.9 | 0.47 |
| Item 16 | 8.1 | 0.27 |
| Item 17 | 7.9 | 0.30 |
| Item 18 | 9.0 | 0.43 |
| Item 19 | 2.7 | 0.31 |
| Item 20 | 10.3 | 0.35 |
| Item 21 | 0.8 | 0.20 |
| Item 22 | 2.0 | 0.23 |
| Item 23 | 12.6 | 0.47 |
| Item 24 | 6.8 | 0.54 |
| Item 25 | 19.5 | 0.27 |
| Item 26 | 8.8 | 0.34 |

Note: Table A1 presents items prompts.

Source: author

The Kaiser-Meyer-Olkin of 0.81 shows that the sample is adequate for conducting factor analysis (Kaiser, 1974). A factor model was fit to determine whether a 1-factor solution could represent the data. As shown in Figure 1, the first factor explained 76 percent of total variance, suggesting that the mental health scale used in the ENSM is capturing a single underlying dimension (i.e., mental health problems). As shown in Figure 2, a summary score following a single factor solution provides a skewed scale, as was expected given the low prevalence of the different items assessed and given the purpose of the measurement (i.e., to identify mental health problems).

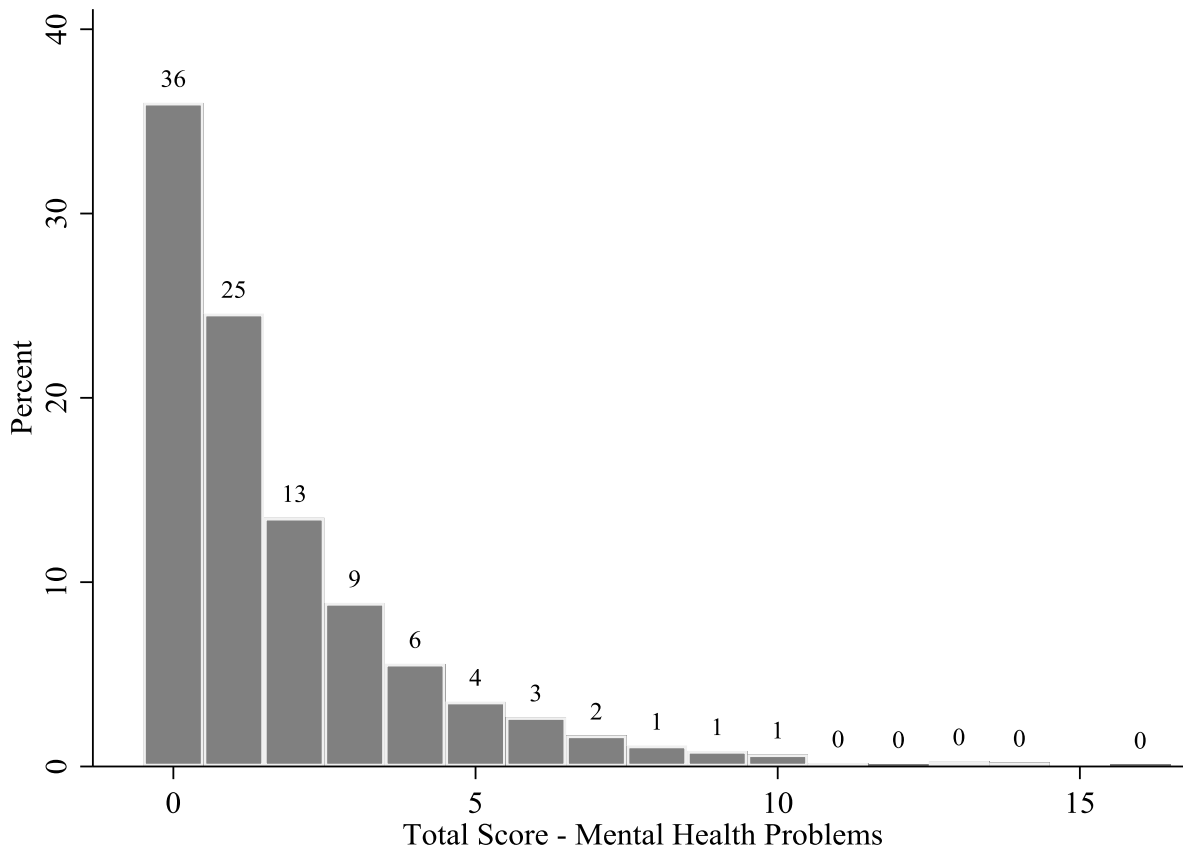
Figure 1. Scree plot of eigenvalues



Note: This figure shows the variation accounted for by each dimension (out of 26) based on a factor analysis of standardized variables ($N = 2,727$)

Source: author

Figure 2. Scale total score – CTT score



Source: author

3.2 IRT and scale information

Table 3 summarizes item discrimination and location parameters for the 2pl-IRT model fitted to the data. Consistent with the findings from CTT, items have high location parameters, ranging from 1.31 (item 12, “Have you noticed that the child has difficulties making friends of his or her same age?”) to 4.71 (item 25, “Is the child eating too little and losing weight?”) standard deviations above the mean. Furthermore, there is a larger dispersion in the estimated discrimination, ranging from 0.3 (item 25) to 2.54 (item 24, “Do you think the child is overeating for his age?”). Nonetheless, in general, most items have *as* parameters above one, suggesting that they distinguish among children with different levels of mental health problems. Figure A2 presents Item Characteristic Curves (ICC), where higher location parameters shift the ICC to the left, and a steeper slope reflects a higher discrimination parameter.

Table 3. Item discrimination and location parameters based on a 2pl-IRT model ($N = 2,727$)

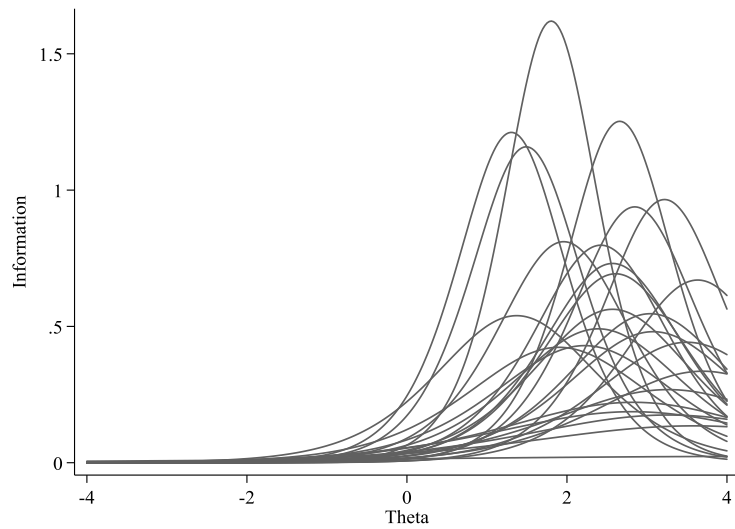
| | Discrimination parameter estimates (a) | Location parameter estimates (b) |
|---------|---|---|
| Item 1 | 1.16 | 3.69 |
| Item 2 | 1.96 | 3.21 |
| Item 3 | 1.78 | 2.42 |
| Item 4 | 1.71 | 2.57 |
| Item 5 | 2.24 | 2.66 |
| Item 6 | 1.93 | 2.85 |
| Item 7 | 2.15 | 1.49 |
| Item 8 | 1.46 | 1.37 |
| Item 9 | 1.40 | 2.39 |
| Item 10 | 1.50 | 2.57 |
| Item 11 | 1.03 | 3.26 |
| Item 12 | 2.20 | 1.31 |
| Item 13 | 1.38 | 3.07 |
| Item 14 | 1.66 | 2.60 |
| Item 15 | 1.80 | 1.96 |
| Item 16 | 0.73 | 3.58 |
| Item 17 | 0.83 | 3.26 |
| Item 18 | 1.31 | 2.22 |
| Item 19 | 1.47 | 3.04 |
| Item 20 | 0.86 | 2.81 |
| Item 21 | 1.63 | 3.63 |
| Item 22 | 1.32 | 3.50 |
| Item 23 | 1.30 | 1.89 |
| Item 24 | 2.54 | 1.80 |
| Item 25 | 0.30 | 4.71 |
| Item 26 | 0.94 | 2.83 |

Note: Figure A2 presents item characteristic curves (ICC).

Source: author

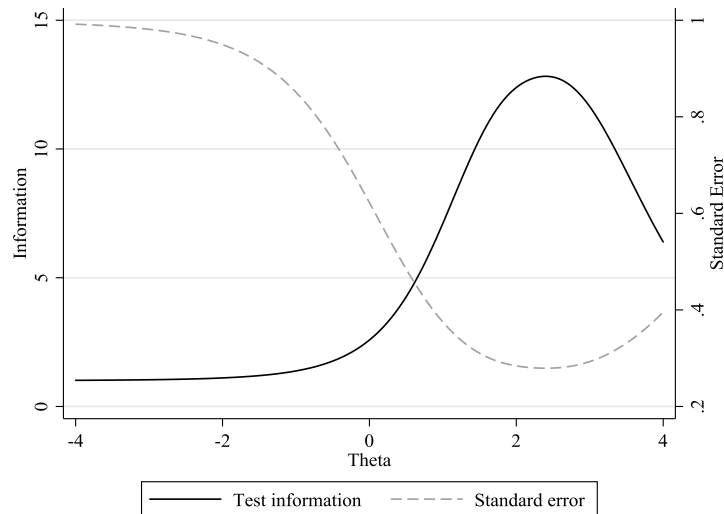
Even though high discrimination parameters are required to distinguish between children with different levels of mental health problems, the information provided by each item depends on its location along the latent scale. In the case of the mental health scale, most items have high location values, indicating that the items are able to distinguish among (and provide information for) children with the presence of mental health problems (which is the purpose of this measurement), as shown in Figure 3. Figure 4 also reflects this fact, showing that the overall scale provides more information for higher levels of theta. Consequently, the scale is more precise and reliable at higher levels of theta, having a lower conditional standard error of measurement between $\theta = 2 - 3$ SD above the mean.

Figure 3. Item information functions (IIF) from a 2pl-IRT model ($N = 2,727$)



Source: author

Figure 4. Test information function and conditional standard error of measurement from a 2pl-IRT model ($N = 2,727$).

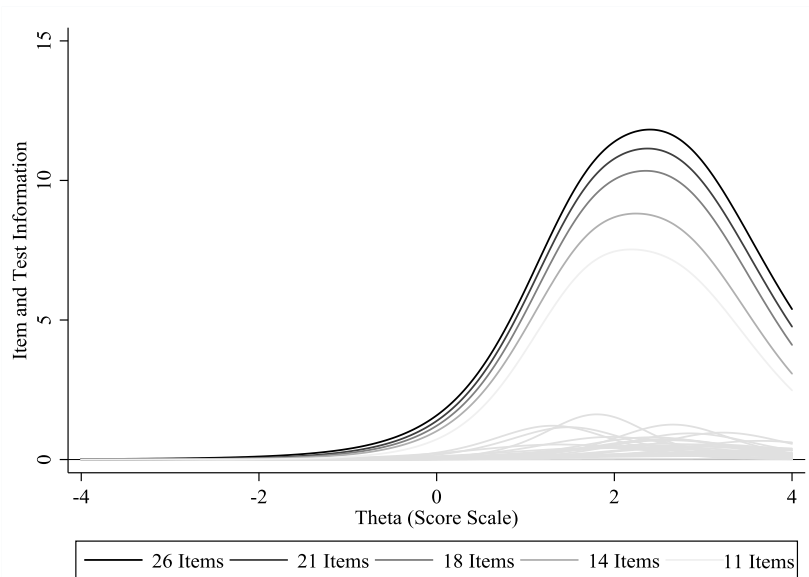


Source: author

IRT main assumption is local independence, which indicates that theta (i.e., the level of the latent trait) provides all the information needed to know the probability of an affirmative response to an item. Given this assumption, it is possible to estimate the information provided by different scales by adding or subtracting the corresponding item information functions (Jessen et al., 2018). Figure 5 presents the test information functions for scales composed by all the 26 items, 21 items, 18 items, 14 items, and 11 items, subtracting items with lower levels of information in a step-wise

fashion. The reduction of items produces lower levels of information, but the reduction is not considerable. In general, even using the 11 most informing items would produce a reliable measure for high levels of theta (particularly around 2 to 3 SD above the mean).

Figure 5. Test information functions from a 2pl-IRT model for different scale specifications ($N = 2,727$)



Source: author

Lastly, the correlational analysis reveals that the mental health scale has a statistically significant association with the expected sign with the number of psychiatric disorders identified using the DISC-IV, as well as with reported physical health, being discriminated, and not being exposed to adversities (Table 4). All these correlations maintain their significance and have similar magnitudes when using reduced forms of the scale, suggesting that a shorter scale may be used if needed given that it is reliable (Figure 5) and has correlational validity evidence (Table 4).

Table 4. Correlation coefficients

| | Convergent validity | | | |
|-----------------------------------|----------------------------|-----------------|---------------|----------------------------|
| | DISC-IV problems | Physical health | Discriminated | Not exposed to adversities |
| 26-item scale ($\alpha = 0.74$) | 0.25** | -0.29** | 0.34** | -0.19** |
| 21-item scale ($\alpha = 0.74$) | 0.25** | -0.27** | 0.34** | -0.18** |
| 18-item scale ($\alpha = 0.73$) | 0.24** | -0.26** | 0.34** | -0.17** |
| 14-item scale ($\alpha = 0.70$) | 0.22** | -0.25** | 0.29** | -0.15** |
| 11-item scale ($\alpha = 0.66$) | 0.23** | -0.25** | 0.29** | -0.15** |

Note: ** $p < 0.01$

Source: author

4. Discussion

In 2015, Colombia undertook its first nationally representative mental health survey (i.e., the ENSM) for children aged between 7 and 11 years. The ENSM included 26 items that were hypothesized to be measuring children's mental health problems, taken from the RQC (Giel et al., 1981), the CBCL (Achenbach, 1999), the CBTD (Caraveo y Anduaga, 2007), and other items were based on the expertise of the group of researchers who designed the survey (Rodriguez et al., 2016). The items were based on measures for which its score interpretation has validity evidence, offering content validity for the ENSM scale. Nonetheless, little was known about validity evidence based on coherence (i.e., internal structure). Moreover, it was not known whether a shorter scale would provide similar levels of information, being more efficient while having a high level of measurement precision.

The purpose of this study was to analyze the psychometric properties of the children's mental health scale using the CTT and IRT frameworks. The findings indicate that the scale has an adequate internal-consistency reliability, and the evidence from factor analysis suggests it is measuring a single latent construct. Furthermore, results from an IRT model shows that most items have a high location as can be expected, reflecting the fact that only individuals with a high θ (i.e., exhibiting mental health problems) will have even or higher odds of answering each item affirmatively. Given the local independence assumption, IRT also reveals that different items provide substantial different levels of information to the scale total information, suggesting a most efficient scale, employing only high-informative items, would be feasible. Indeed, findings from the item information function and convergent validity indicate that shorter scales will keep desirable psychometric properties and could be employed in future implementations of the mental health scale when increased efficiency is needed.

One major strength of this study and contribution to the Colombian literature is the use of an IRT framework, which conversely to CTT theoretically offers population-invariant parameters that can accurately inform future implementations of each item. Indeed, CTT framework produces parameters that depend on the specific population where the items are implemented, and which are, to a certain extent, predictable. For instance, following Spearman-Brown prophecy formula, it is possible to infer that Cronbach's alpha will be higher as one test has more items and as the population where the test is implemented is more heterogeneous, whereas less items and a more

homogenous population would lead to lower test reliabilities (Traub & Rowley, 1991). On the other hand, IRT estimates item location, discrimination, and information that are assumed to hold in different occasions and populations (Embretson & Reise, 2013). These population-invariant parameters can inform the design of scales, maximizing precision at the desired θ level and permitting the selection of high-informing items when efficiency is paramount (Jessen et al., 2018).

Even though this study does relevant contributions offering validity evidence based on coherence and correlation for the ENSM children's mental health scale, it does not provide validity evidence based on response process (i.e., cognition) or consequences (Koretz, 2008). A future pilot study could be implemented to analyze the type of cognitive process respondents employ to respond to the scale's items, assessing whether there are some items that may be particularly cognitive-demanding. Moreover, it could be useful to conduct studies that allow to elucidate whether the test produces certain consequences on respondents, such as changes in their behaviors or interactions with their children following the test.

5. Conclusion

The 26-item children's mental health scale used in the ENSM has adequate psychometric properties, and evidence from factor analysis suggests it is measuring a single latent construct. A 2pl-IRT model reveals that the scale is accurate at distinguishing between children with high levels of θ , around 1 and 3 SD above the mean. In a future implementation of the scale, when lowering the number of items and higher efficiency is required, a 21, 18, 14, and even 11- item scale, may hold desirable properties and predictive power. These findings suggest that future efforts can be conducted to continue monitoring children's mental health in Colombia, especially in the post-conflict situation when it is needed to identify children who would need additional supports.

6. References

- Aarons, G. A., Monn, A. R., Leslie, L. K., Garland, A. F., Lugo, L., Hough, R. L., & Brown, S. A. (2008). Association between mental and physical health problems in high-risk adolescents: a longitudinal study. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, 43(3), 260-267. doi:10.1016/j.jadohealth.2008.01.013
- Achenbach, T. M. (1999). The Child Behavior Checklist and related instruments. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 429-466). NJ, US: Lawrence Erlbaum Associates Publishers.
- Ahmad, A., Abdul-Majeed, A. M., Siddiq, A. A., Jabar, F., Qahar, J., Rasheed, J., & von Knorring, A.-L. (2007). Reporting Questionnaire for Children as a Screening Instrument for Child Mental Health Problems in Iraqi Kurdistan. *Transcultural Psychiatry*, 44(1), 5-26. doi:10.1177/1363461507074949
- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7), 476-487. doi:<https://doi.org/10.1016/j.jpubeco.2010.11.006>
- Anda, R. F., Felitti, V. J., Bremner, J. D., Walker, J. D., Whitfield, C., Perry, B. D., . . . Giles, W. H. (2006). The enduring effects of abuse and related adverse experiences in childhood. *European Archives of Psychiatry and Clinical Neuroscience*, 256(3), 174-186. doi:10.1007/s00406-005-0624-4
- Bell, C. C. (1994). Dsm-iv: Diagnostic and statistical manual of mental disorders. *JAMA*, 272(10), 828-829. doi:10.1001/jama.1994.03520100096046
- Brennan, R. L., National Council on Measurement in Education, & American Council on Education. (2006). *Educational measurement* (4th ed. ed.). Westport, CT.
- Caraveo y Anduaga, J. J. (2007). Cuestionario breve de tamizaje y diagnóstico de problemas de salud mental en niños y adolescentes: algoritmos para síndromes y su prevalencia en la ciudad de México. Segunda parte. *Salud Mental*, 30(1), 48-55.
- Castro, J., Billick, S. B., & Swank, A. C. (2016). Utility of a New Spanish RQC and PSC in Screening with CBCL Validation. *Psychiatric Quarterly*, 87(2), 343-353. doi:10.1007/s11126-015-9391-1

- Cooke, C. L., Bowie, B. H., & Carrère, S. (2014). Perceived Discrimination and Children's Mental Health Symptoms. *Advances in Nursing Science*, 37(4), 299-314. doi:10.1097/ans.0000000000000047
- Crocker, L. M., & Algina, J. (1986). Introduction to classical and modern test theory.
- Dimitry, L. (2012). A systematic review on the mental health of children and adolescents in areas of armed conflict in the Middle East. *Child: Care, Health and Development*, 38(2), 153-161. doi:doi:10.1111/j.1365-2214.2011.01246.x
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory*: Psychology Press.
- Erskine, H. E., Baxter, A. J., Patton, G., Moffitt, T. E., Patel, V., Whiteford, H. A., & Scott, J. G. (2017). The global coverage of prevalence data for mental disorders in children and adolescents. *Epidemiology and Psychiatric Sciences*, 26(4), 395-402. doi:10.1017/S2045796015001158
- Giel, R., de Arango, M. V., Climent, C. E., Harding, T. W., Ibrahim, H. H. A., Ladrado-Ignacio, L., . . . Younis, V. O. A. (1981). Childhood Mental Disorders in Primary Health Care: Results of Observations in Four Developing Countries. *Pediatrics*, 68(5), 677-683.
- Jessen, A., Ho, A. D., Corrales, C. E., Yueh, B., & Shin, J. J. (2018). Improving Measurement Efficiency of the Inner EAR Scale with Item Response Theory. *Otolaryngology–Head and Neck Surgery*, 158(6), 1093-1100. doi:10.1177/0194599818760528
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36. doi:10.1007/bf02291575
- Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., . . . Rahman, A. (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet*, 378(9801), 1515-1525. doi:[https://doi.org/10.1016/S0140-6736\(11\)60827-1](https://doi.org/10.1016/S0140-6736(11)60827-1)
- Koretz, D. M. (2008). *Measuring up: what educational testing really tells us*. Cambridge, Mass.
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley.
- Ministerio de Salud & Colciencias (2015). Encuesta nacional de salud mental 2015. In. Bogotá: Ministerio de Salud & Colciencias.

- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1-18. doi:[https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- OIM, UNICEF, & ICBF. (2013). *Estado psicosocial de los niños, niñas y adolescentes: una investigación de consecuencias, impactos y afectaciones por hecho victimizante con enfoque diferencial en el contexto del conflicto armado colombiano*.
- Red Nacional de Información (2018). Registro Único de Víctimas (RUV). Retrieved from <https://www.unidadvictimas.gov.co/es/registro-unico-de-victimas-ruv/37394>
- Rodríguez, N., Rodríguez, V. A., Ramírez, E., Cediell, S., Gil, F., & Rondón, M. A. (2016). Aspectos metodológicos del diseño de muestra para la Encuesta Nacional de Salud Mental 2015. *Revista Colombiana de Psiquiatría*, 45, 26-30. doi:<https://doi.org/10.1016/j.rcp.2016.08.009>
- Rodríguez, V., Moreno, S., Camacho, J., Gómez-Restrepo, C., de Santacruz, C., Rodríguez, M. N., & Tamayo Martínez, N. (2016). Diseño e implementación de los instrumentos de recolección de la Encuesta Nacional de Salud Mental Colombia 2015. *Revista Colombiana de Psiquiatría*, 45, 9-18. doi:<https://doi.org/10.1016/j.rcp.2016.10.001>
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, Differences From Previous Versions, and Reliability of Some Common Diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(1), 28-38. doi:10.1097/00004583-200001000-00014
- StataCorp (2017). Stata statistical software: release 15. College Station, TX: StataCorp LLC.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC, US: American Psychological Association.
- Traub, R. E., & Rowley, G. L. (1991). Understanding Reliability. *Educational Measurement: Issues and Practice*, 10(1), 37-45. doi:doi:10.1111/j.1745-3992.1991.tb00183.x

7. Appendix

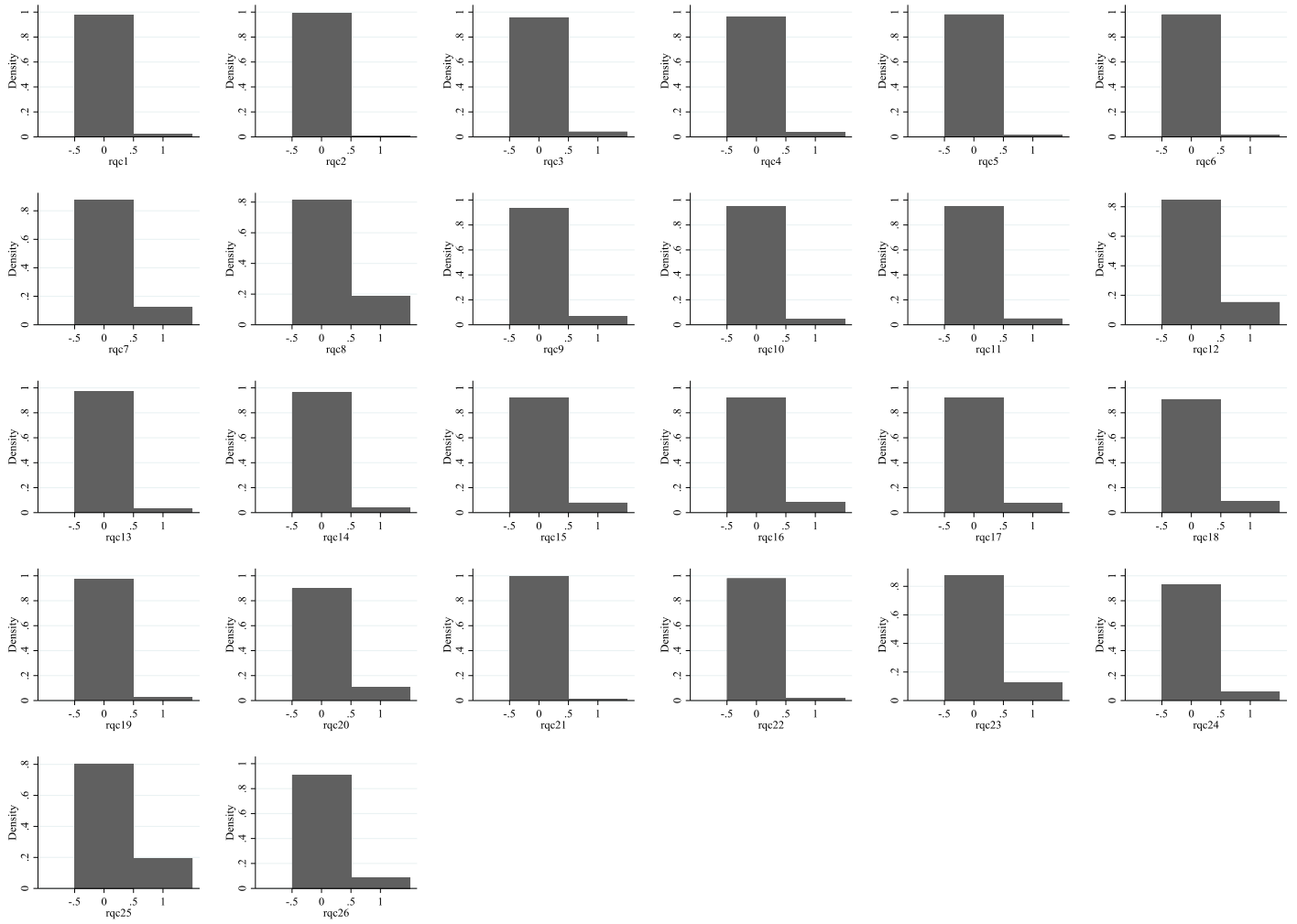
Table A1. Item prompts

| Item | Label | Label in original language |
|---------|---|--|
| Item 1 | Is the child speech in any way abnormal? | ¿El lenguaje del niño es anormal en alguna forma? |
| Item 2 | Does the child sleep badly? | ¿El niño duerme mal? |
| Item 3 | Did the child ever have a fit or fall to the ground for no reason? | ¿Ha tenido el niño en algunas ocasiones convulsiones o caídas al suelo sin razón? |
| Item 4 | Does the child suffer from frequent headaches? | ¿Sufre el niño de dolores fuertes de cabeza? |
| Item 5 | Does the child run away from home frequently? | ¿El niño ha huido de la casa frecuentemente? |
| Item 6 | Does the child steal things from home? | ¿Ha robado cosas de la casa? |
| Item 7 | Does the child get scared or nervous for no good reason? | ¿Se asusta o pone nervioso sin razón? |
| Item 8 | Does the child in any way appear backward or slow to learn as compared with other children of about the same age? | ¿Parece como retardado o lento para aprender? |
| Item 9 | Does the child nearly never play with other children? | ¿El niño casi nunca juega con otros niños? |
| Item 10 | Does the child wet or soil himself/herself? | ¿El niño se orina o defeca en la ropa? |
| Item 11 | Has the child stopped talking seasonally or at all? | ¿El niño ha dejado de hablar por temporadas o del todo? |
| Item 12 | Have you noticed that the child has difficulty making friends of the same age? | ¿Ha notado que al niño se le dificulte hacer amigos de su misma edad? |
| Item 13 | Does the child tend to be alone? | ¿El niño tiende a permanecer solo? |
| Item 14 | Does the child frequently walk with difficulty or accidentally hit himself? | ¿El niño frecuentemente camina con dificultad o se golpea accidentalmente? |
| Item 15 | Does the child exhibit strange behaviors such as talking alone without playing? | ¿El niño presenta comportamientos extraños como hablar solo sin estar jugando? |
| Item 16 | Has the child had trouble learning to read or write? | ¿El niño ha tenido problemas para aprender a leer o escribir? |
| Item 17 | Has the child had trouble learning math? | ¿El niño ha tenido problemas para aprender matemáticas? |
| Item 18 | Has the child repeatedly been a victim of abuse or physical or psychological maltreatment? | ¿El niño, repetidamente, ha sido víctima de abuso o maltrato, físico o psicológico? |
| Item 19 | Has the child molested or repeatedly assaulted other children? | ¿El niño ha molestado o ha agredido repetidamente a otros niños? |
| Item 20 | Can't concentrate, can't pay attention for long? | ¿El niño ha tenido problemas para fijar y mantener la atención o concentrarse? |
| Item 21 | Has the child needed to change school more than 3 times? | ¿El niño ha necesitado cambio de institución escolar más de 3 veces? |
| Item 22 | Does the child refuse to go to school repeatedly? | ¿El niño se niega a ir a la escuela repetidamente? |
| Item 23 | Does the child have difficulty following rules, limits or respecting authority figures? | ¿El niño tiene dificultad para seguir normas, límites o respetar figuras de autoridad? |
| Item 24 | Do you think the child is overeating for his age? | ¿Considera que el niño está comiendo en exceso para su edad? |

| | | |
|---------|---|---|
| Item 25 | Is the child eating too little and losing weight? | ¿El niño está comiendo muy poco y ha bajado de peso? |
| Item 26 | Does the child repeatedly complain of pain, dizziness, and desire to vomit or other ailments without medical explanation? | ¿El niño se queja repetidamente de dolores de estómago, extremidades, de mareos, ganas de vomitar u otras dolencias sin explicación médica? |

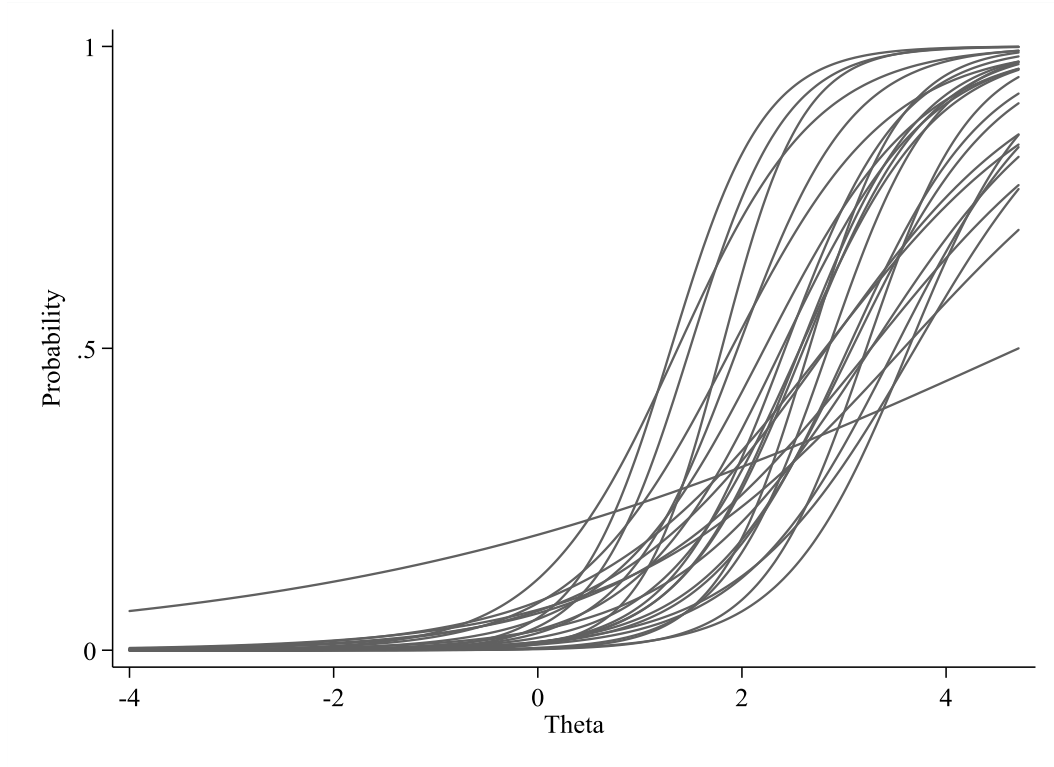
Source: author

Figure A1. Item histograms



Source: author

Figure A2. Item characteristic curves (ICC)



Source: author



NUEVA

Maestría en Gestión Pública



Título otorgado

Magíster en Gestión Pública

SNIES

106656 Registro calificado: resolución No. 572 del 22 de enero de 2018, por 7 años

Duración

Un año y medio (36 créditos académicos, distribuidos en 3 semestres)

Modalidad

Presencial en Bogotá

Dirigido a

Profesionales con más de dos años de experiencia laboral vinculados al sector público, privado, organizaciones no gubernamentales y sin ánimo de lucro

La Maestría se ofrecerá en sesiones presenciales y virtuales, que permitirá cursarla desde cualquier lugar del país

Maestría en Políticas Públicas



Título otorgado

Magíster en Políticas Públicas

SNIES

90798 Registro calificado: resolución No. 2056 del 17 de febrero de 2015, por 7 años

Duración

2 años (42 créditos académicos, distribuidos en cuatro semestres)

Modalidad

Presencial en Bogotá

Dirigido a

Profesionales sobresalientes que demuestren interés por los asuntos públicos

Maestría en Salud Pública



Título otorgado

Magíster en Salud Pública

SNIES

91281 Registro calificado: resolución No. 20781 del 9 de octubre de 2017, por 7 años

Duración

2 años (44 créditos académicos, distribuidos en cuatro semestres)

Modalidad

Presencial en Bogotá

Dirigido a

Profesionales de diversas disciplinas con interés en aportar a la discusión, el análisis, el diseño, la implementación y la evaluación de las políticas en salud pública

Más información



Escuela de Gobierno Alberto Lleras Camargo
Universidad de los Andes
Carrera 1 No. 19 - 27 - Bloque AUJ tercer piso
teléfono: 3394949 ext. 2073 - Bogotá, Colombia



egob.uniandes.edu.co



fb.com/EGOBUniandes



[@EGOBUniandes](https://twitter.com/EGOBUniandes)

Documentos de trabajo EGOB es una publicación periódica de la Escuela de Gobierno Alberto Lleras Camargo de la Universidad de los Andes, que tiene como objetivo la difusión de investigaciones en curso relacionadas con asuntos públicos de diversa índole. Los trabajos que se incluyen en la serie se caracterizan por su interdisciplinariedad y la rigurosidad de su análisis, y pretenden fortalecer el diálogo entre la comunidad académica y los sectores encargados del diseño, la aplicación y la formulación de políticas públicas.

egob.uniandes.edu.co

 fb.com/EGOBuniandes  [@EGOBUniandes](https://twitter.com/EGOBUniandes)